Experiments with Geiger-Müller Counters and Associated Circuits

A. L. HUGHES

Wayman Crow Hall of Physics, Washington University, St. Louis, Missouri

DURING the last few years, numerous papers have been written on Geiger-Müller counters and the circuits associated with them. In every case the purpose has been to describe an improvement in the use of the counter as a research tool, or an investigation of the way it operates. The present paper is written for a different purpose. It seemed desirable to plan a group of experiments on Geiger-Müller counters as part of an advanced laboratory course designed to provide graduate students with a variety of skills in experimental methods. The Geiger-Müller counter has now become such an important tool that a student should certainly have an opportunity to gain some degree of familiarity with the way it is used, even though his principal research problem may not involve counters at all. An equally important reason for experiments on counters is the following. The standard laboratory experiments, both elementary and advanced, emphasize exact reproducibility of observations as a criterion of good experimental work; they give the student no suggestion that many physical phenomena are essentially statistical in nature. In almost every laboratory problem on which the student has had to work, it has been tacitly assumed that any deviations between the values he finds and the "true" values must be attributed either to his inaccuracy or to instrumental defects. Under such circumstances it is natural to conclude that, given infinite skill and patience, all physical

measurements when made under the same conditions should always give precisely the same results. Experiments with Geiger-Müller counters are especially valuable in correcting this situation, for with them statistical fluctuations can be readily observed and investigated.

The equipment and experiments described in this paper are planned to acquaint the student with phenomena in which the outstanding feature is their statistical character, to show how observations on such phenomena may be made and interpreted, and to give familiarity with Geiger-Müller counters and their associated circuits. The purpose is not to describe a new circuit or a new way of handling observations, but to describe those circuits which seem to be most suitable in a first approach to an experimental study of counters and to show what kind of observations can be made with them and to what conclusions they lead. Inasmuch as these studies on G-M counters are but a small part of a general course on advanced laboratory methods, it was deemed necessary to limit the experiments to those which would provide the most useful information and technics. A self-imposed limitation is that not more than 5 laboratory periods (approximately 20 hours) should be set aside for the study of G-M counters. This will account for the fact that many obviously desirable experiments—for example, on "coincidences"— are omitted.

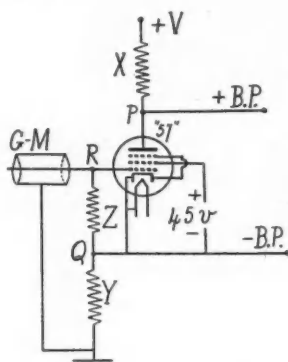Before a final decision was made as to the

FIG. 1. Schematic diagram of Neher-Pickering control circuit. $X = 200,000$ ohms; $Y = 1$ megohm; $Z = 5$ megohms.

most suitable type of equipment to assemble, a careful experimental study was made of several of the various control, scaling and recording circuits which have been described in the literature. The criterions used in selecting the most suitable circuits included simplicity, dependability, tolerance to reasonable variations in conditions, as well as instructional value. Several of the published scaling circuits were found to be quite difficult to adjust, and the slightest variations in conditions (such as drift in the grid voltage) would make the circuit inoperative. The Reich circuit, which was finally adopted as the scaling circuit, was exceptionally good, for wide variations in operating conditions did not interfere with normal scaling. The Neher-Pickering control circuit, the Reich scaling circuit and the Pickering recording circuit were finally chosen as the most suitable for our special requirements.

A typical G-M unit designed to be used as a tool for research is usually made as compact as possible; once the unit has been properly adjusted, accessibility of the various elements is of secondary importance. In our unit, on the other hand, the arrangement of the various circuits was deliberately planned so that all the elements are readily accessible. By temporarily connecting voltmeters and milliammeters to suitable points—voltages across and currents through—most of the elements can readily be measured. The accessibility of the elements in the circuit makes it possible, with very little trouble, to observe the effect of changing any resistor or condenser. While it is customary, because of convenience, to have G-M units

entirely "a.c.-operated," we have not hesitated to use B-batteries where it seemed more desirable from an instructional point of view.

In Part II, an account of that part of statistics bearing on counters and scaling circuits is given. The results are of course not new. It is believed that the method of deriving them will be welcomed by some readers as an easily understood alternative to the more elegant and condensed methods given elsewhere.

I. EXPERIMENTAL

A. Control circuit

The functions of the control circuit are (1) to quench the discharge initiated in the G-M tube by the passage of an ionizing particle through it and (2) to pass on a suitable electric pulse to the recording device, either directly or through a scaling circuit. The Neher-Pickering control circuit[1] was selected as being suitable for our purpose.

1. *Operating principle.*—The Neher-Pickering circuit is shown in Fig. 1. When there is no discharge through the G-M tube, the control grid of the 57 tube takes up a potential which is slightly negative with respect to the cathode so that there is a small current from $+V$ through the tube to ground. The potential of the anode wire of the G-M tube is nearly equal to that of the point Q. By choosing $+V$ properly, the voltage across the G-M tube can be made such that when an ionizing particle passes through the tube a discharge occurs. When this happens, the potential of R becomes more negative than before and so stops the current through the 57 tube, with the result that the potential of P (the plate) rises towards $+V$ while the potential of Q (the cathode) falls towards that of the ground. But as the potential of Q, and therefore that of R, drops, the potential across the G-M tube falls below the value necessary to maintain the discharge. This quenching of the discharge allows the potential of R with respect to that of the cathode to regain its original value, with the result that the tube becomes conducting once more and ready to respond to another ionizing

[1] H. V. Neher and W. H. Pickering, Phys. Rev. **53**, 316 (1938). See also J. Strong, *Procedures in Experimental Physics* (Prentice-Hall), p. 279.

particle. The cycle of events is completed within a very short time.[2] It will be noticed that the momentary changes in the potentials of P and Q are such that a positive electric pulse appears at P and a negative one at Q. Either one can be passed to the next circuit through a condenser.

2. *Circuit layout.*—The control circuit, for which Fig. 1 is a schematic drawing, is laid out on a board 10×12 in., as in Fig. 2, so that every part is accessible. Voltages across any element in the circuit can readily be measured. For such measurements, a vacuum tube voltmeter is used, as ordinary voltmeters draw too much current and disturb the potential distribution. Any resistor may easily be replaced by one of different value, should it be desired to study the effect of such a change. (The resistors are 1-w metallized IRC resistors.) The high voltage to be applied at $+V$ is provided by a conventional half-wave rectifier,[3] and measured by an inverse vacuum tube voltmeter of the type designed by Schmitt.[4] The filament transformer must have its coils well insulated from each other since the heater of the 57 tube in the control circuit may be as much as 1000 v above ground.
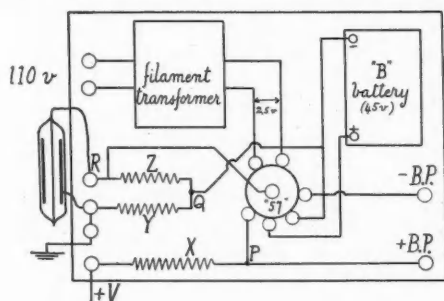
A Pickering recording circuit requires a posi-

FIG. 2. Neher-Pickering control circuit, showing relative locations of component parts.

tive pulse to operate it, whereas a Reich scaling circuit operates only on a negative pulse. Two binding posts, providing positive and negative pulses, respectively, are shown at the right-hand side of Fig. 2.

The G-M tube, made according to the technic of Locher, was obtained from the Herbach and Rademan Company, of Philadelphia. While experience in the construction of G-M tubes has undoubted value, it was considered better to start with a good G-M tube and to devote all the available time to a study of its characteristics and those of the associated circuits.[5]

3. *Characteristics.*—(*a*) In one set of measurements, when the high voltage unit supplied 990 v to the control circuit (that is, at $+V$), the potentials at P and Q were 825 and 815 v, respectively. This refers to the situation when there is no discharge through the G-M tube. It is impossible to measure, with any kind of voltmeter, the distribution of voltages during the passage of a momentary discharge through the G-M tube. An oscillograph would be necessary. The steady current (0.7 ma) through the 57 tube could be measured by a milliammeter inserted between Y and the ground.

(*b*) The foregoing description covers the observations that can be easily made on the control circuit itself. It is convenient, however, to take up at this point an examination of the operating range of the G-M tube, or its "plateau." The binding post of the control circuit at which positive pulses appear is connected to the recording circuit (described in Section **B**). A suitable radioactive source[6] is placed near the G-M tube and the voltage applied to the control circuit is increased until the counting just begins. This gives the threshold of the G-M tube and control circuit combined. The threshold of the G-M tube itself is less than this, and may be calculated approximately by multiplying it by the ratio $Y/(X+Y)$, where Y and X are the resistances shown in Figs. 1 and 2. The next step is to measure the counting rate as a function

---

[2] Measurements made by Lifschutz and Duffendack [Phys. Rev. **54**, 714 (1938)] with a Neher-Harper control circuit give values of the order 10^{-3} to 10^{-4} sec. The same order of magnitude may be expected with a Neher-Pickering control circuit.

[3] The rectifier is similar to that shown in Strong's *Procedures in Experimental Physics*, p. 292. A fair degree of regulation is provided by a ballast tube in the 110-v supply line, together with a suitable resistance in parallel with the primary of the transformer to insure that the current through the ballast tube has the value for best regulation. The greater regulation provided by a properly designed voltage regulator, such as those shown on pp. 294–296 of the same book, would, of course, be better.

[4] O. H. A. Schmitt, J. Sci. Inst. **15**, 136 (1938).

[5] Satisfactory tubes have been made by us, following the technic of Duffendack, Lifschutz and Slawsky, Phys. Rev. **52**, 1231 (1937).

[6] Old radon tubes form very convenient sources. They may be cemented by Glyptal or Duco cement to the bottom of a very shallow depression in a brass plate $3 \times 1 \times \frac{1}{8}$ in. Several such plates with 1, 5, 20 and 100 tubes, respectively, provide a useful assortment of sources.
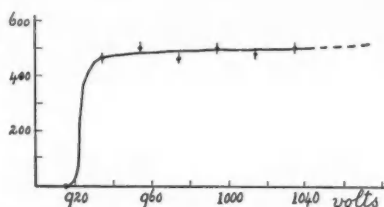
FIG. 3. "Plateau" of the G-M tube. Ordinates are total counts in a 10-min interval. Abscissas are voltages applied to the control circuit. Voltages across the G-M tube are approximately 0.83 of these.

of the applied voltage. It is convenient to increase the voltage in steps of about 30 v, and to place the radioactive source at such a distance that the counting rate is of the order of 30 per min. Fig. 3 shows that the plateau—that is, the region over which the counting rate does not change much with the voltage—is at least 100 v wide. The experimental values are represented by short vertical lines, rather than by points, in recognition of the fact that repetitions of any one observation involving a finite number of counts cannot, because of statistical fluctuations, be expected to give identical values. According to statistics, the probable error of the number of pulses during any one interval when they arrive at random is $0.67n^{\frac{1}{2}}$, where n is the average number arriving during that interval. This means that if the interval is such that, on the average, 100 pulses arrive, the number arriving during any one interval is as likely to be between the limits 93 and 107 as outside these limits.

B. Recording Circuit

1. *Operating principle.*—The recording circuit used is a simple, but effective, circuit devised by Pickering.[7] It is shown in Fig. 4. The tube is the "885" argon-filled Thyratron. When the tube is nonconducting, which is its normal state, the potential of P, the cathode, is determined by V and the ratio R_3/R_2 (disregarding the resistance of the counter CC and R_1, which are relatively small). The grid D, which is at ground except when a pulse arrives, is sufficiently negative with respect to the cathode P to keep the tube in a nonconducting state. If a pulse of sufficient magnitude is applied at B, the grid

[7] W. H. Pickering, Rev. Sci. Inst. **9**, 180 (1938).

becomes sufficiently less negative (or even positive) so that the Thyratron is made conducting and acts as though it had little resistance. The potential of P rises rapidly and approaches that of the plate. At the same time, a definite charge, determined by the capacitance of the condenser C, passes from the voltage source through the mechanical counter CC, causing it to register one count. The Thyratron discharge stops almost immediately because the potential of the grid becomes highly negative with respect to that of the cathode, while at the same time the voltage difference between the plate and cathode has fallen to a small value. (Ordinarily the grid in a Thyratron loses control when once the Thyratron fires, but if the Thyratron current is not too large and if the potential between the plate and cathode is reduced to a small value, as is the case here, a sufficiently large negative potential on the grid enables it to regain control.)

The heater of the Thyratron in this circuit must be supplied with current from a transformer whose secondary is *not* grounded, in order that the potential of the cathode may ride up and down freely.

The capacitance of the condenser C need not be larger than is necessary to trip the mechanical counter each time a pulse arrives. Too large a capacitance may result in a diminution of the maximum counting rate by transferring the location of the principal source of losses in counting from the mechanical counter to the electrical part of the recording unit.

2. *Circuit layout.*—The recording circuit was mounted on a board 8×6 in. as shown in Fig. 5. This arrangement makes it possible to change any element readily and see what is the result of that change. The pulse, which must be a positive pulse, may be fed in at B through a
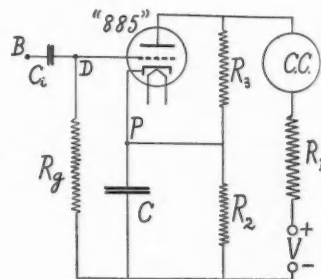
FIG. 4. Schematic diagram of Pickering recording circuit.

50-$\mu\mu$f condenser. It is important to notice that neither the cathode nor the heater is grounded. (The center tap of the low voltage side of the filament transformer, which in many applications is grounded, is here left unconnected.) The reason is given in Section **B**,1. The voltage is supplied by a conventional power pack providing any desired voltage between 200 and 500 v. A Cenco high impedance counter or a Western Electric message register (type 12G) may be used. A larger condenser at C is necessary to operate the latter than to operate the former.

3. *Tests on the recording circuit.*—(*a*) In order to study the characteristics of the recording and scaling circuits it is very convenient to have a device for producing pulses whose maximum voltage can be varied at will. A simple form is shown in Fig. 6. The closing of the switch causes a pulse to appear on the other side of the condenser. It was found that snap switches, tapping keys, etc., were unsatisfactory, especially for testing the scaling circuit, because the act of closing the circuit frequently caused two or more pulses to appear in close succession. This can be avoided by the arrangement shown in Fig. 7. A platinum wire, 5 mils in diameter, welded to the end of a hinged lever, can be driven rapidly into the mercury by the flick of a finger. When very clean mercury is used, and when the platinum wire has been cleaned by passing a flame over it, the device gives single pulses.

(*b*) It is of interest to find the range of applied voltages over which the recording circuit will operate. For example, with $R_1 = 2000$ ohms,
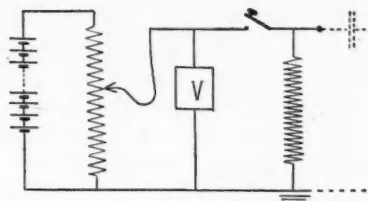

FIG. 5. Pickering recording circuit, showing relative locations of parts.

FIG. 6. Diagram of device for producing pulses

$R_3 = 200,000$ ohms, $R_2 = 25,000$ ohms, $C = 0.25$ μf, $C_i = 50$ $\mu\mu$f and $R_g = 100,000$ ohms, the device in Fig. 6 was used to determine the minimum pulse-voltage necessary to cause the recording circuit to operate the Cenco counter, as a function of the applied voltage V. With V set at 320, 280, 250 and 237 v, the recording circuit would trip with minimum pulse-voltages of 10.0, 8.5, 7.0, and 8.5 v, respectively. Any pulse-voltage *above* this, up to the highest tried (60 v) would trip the circuit. Below $V = 225$ v, no pulse could be found that would trip the circuit. Therefore the best value for V is between 250 and 270 v. It is important to note that the circuit is not critical; it will operate over a wide range of values of V.

(*c*) With the constants of the circuit as given in (*b*), a point worth investigating is the effect of the value of the ratio R_3/R_2 on the behavior of the circuit. In this particular circuit, with $R_3 = 200,000$ ohms, R_2 was varied. When R_2 was as low as 20,000 ohms, occasional double counts appeared for single pulses. When it was between 23,000 and 30,000 ohms, normal single counts appeared for single pulses. As it approached the 30,000-ohm value the required minimum input pulse increased considerably. This suggests that $R_2 = 25,000$ ohms is a suitable value to use permanently.

(*d*) A Cenco high impedance counter is found to respond with $C = 0.25$ μf. On replacing it by a Western Electric message register (type 12G) the capacitance must be increased to 1.75 μf. Since we normally use the Cenco counter with our recording circuit, it was found convenient to connect temporarily extra condensers in parallel with the 0.25-μf condenser already on the board, when investigating the changes necessitated by the substitution of the message register for the Cenco counter.
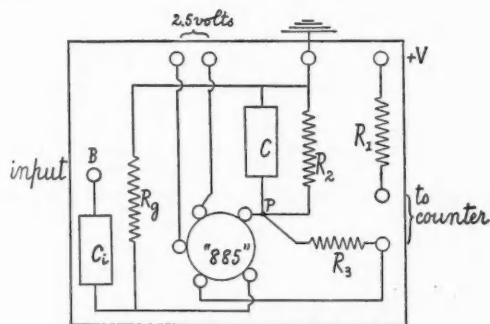
C. Scaling circuit

A scaling circuit is often valuable when the pulses delivered by the control circuit to the recording circuit follow one another so rapidly that the recording circuit fails to record many of the pulses. The scaling circuit is a device so constructed that, when it receives pulses from the control circuit, it delivers every 2nd, 3rd, 4th, $\cdots$ pulse to the recording circuit. Thus the average time between the pulses received by the recording circuit is increased. Also, it can be shown that the number of very short pulses—which are the ones missed by the mechanical counter—is greatly reduced. The earliest scaling circuit was designed by Wynn-Williams.[8] In this type of scaling circuit, the scaling factor is 2^n, where n is the number of stages. Most, but not all, of the later scaling circuits work on the same general principle, namely, that each stage hands on to the next every alternate pulse which it receives.

Before deciding on which scaling circuit to adopt for use in this study several well-known scaling circuits were tried. In most cases it was found that the adjustment needed considerable care and that the tolerances were quite narrow. By this is meant that a small change in some working condition (for example, in a grid bias) would put the scaling circuit out of action. The scaling circuit due to Reich[9] was finally adopted because it combines unusual simplicity with an ability to tolerate large changes in operating conditions without any effect whatever on its scaling properties.

1. *Operating principle.*—A single stage in the Reich scaling circuit is shown in Fig. 8. A characteristic of this circuit is that it has two
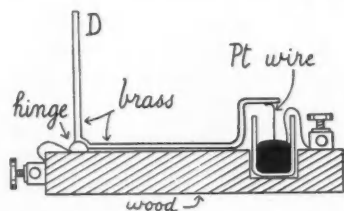

Fig. 7. Construction of the switch used in Fig. 6.

[8] C. E. Wynn-Williams, Proc. Roy. Soc. **136**, 321 (1932).
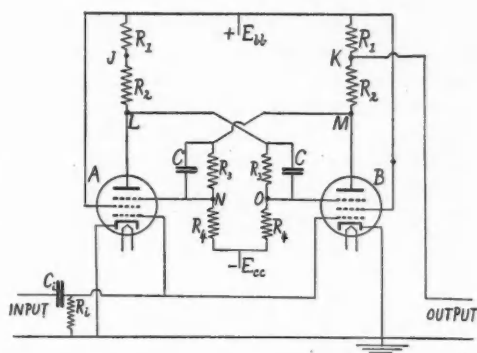[9] H. J. Reich, Rev. Sci. Inst. **9**, 222 (1938).

Fig. 8. Schematic diagram of one stage in the Reich scaling circuit.

stable conditions. Either tube A is conducting and tube B is nonconducting, or the reverse is the case. If a suitable *negative* pulse is fed in, the circuit "trips;" that is, the conducting tube becomes nonconducting, and vice versa. If a *positive* pulse is applied to the input, nothing happens. The circuit therefore responds only to negative pulses and thus passes on only every second pulse, if the pulses it receives are alternately positive and negative.

To explain how the circuit trips we shall list the voltages which were found to exist at certain points when tube A was conducting and tube B nonconducting, and when $E_{bb}=45$ v and $E_{cc} = -28$ v. Tube B is nonconducting because its suppressor grid is at a high negative voltage, -12.5 v.

Let us suppose that a *positive* pulse is applied through the input condenser C_i to the control grids of both tubes. Both tubes are thereby made actually or potentially more conducting, but this does not change the situation in tube B because the negative potential of its suppressor grid keeps it in a nonconducting state. Hence tubes A and B do not interchange roles; that is,

TABLE 1. *Voltage distribution in the Reich scaling circuit.*

CONDUCTING TUBE		NONCONDUCTING TUBE	
LOCATION	VOLTS	LOCATION	VOLTS
J	$+37.0$	K	$+42.5$
L	$+ 2.0$	M	$+35.0$
N	$+ 2.3$	O	-12.5

there is no tripping. The effect of a *negative* pulse is very different. As the control grid of tube A becomes more negative than before, the current through it diminishes, and the potentials of L and O rise from their values of $+2.0$ and -12.5 v, respectively. If the potential of O is raised enough to make tube B slightly conducting (it does not have to go the whole way from -12.5 to 0), the potential of M, and therefore of N, drops. This makes the suppressor grid of tube A more negative than it was. The result is a further drop in the current through tube A combined with a further rise in the potential of L, which in turn raises the potential of the suppressor grid of tube B and makes it more conducting. At the same time the potential of the suppressor grid of tube A is driven to a negative value and the current through it stops. The tripping of the circuit takes place in a very short time, probably less than 10^{-5} sec.

The function of the condensers C is to transmit very rapidly the changes of potentials from the points L and M to the corresponding suppressor grids. These are then held at their respective potentials through the resistances R_3.

The potential of the point K moves up or down each time the current trips. This is applied to the next stage through a suitable condenser. The second stage is tripped just half as often as the first stage.

2. *Circuit layout.*—In order to be able to make measurements on the various elements in the scaling circuit, it is necessary to plan its arrangement so that every element is easily reached. Accessibility can be obtained at the price of compactness, which is unimportant in our application. A three-stage Reich scaling circuit was assembled on a board 28×12 in. The underside was covered with galvanized iron which was grounded. When necessary, shielding could be secured by placing a grounded sheet metal box $28 \times 12 \times 7$ in. high over the scaling circuit. Each stage occupied a region 8 in. long on the board.

The arrangement of elements in one stage, viewed from above, is shown roughly to scale in Fig. 9, which corresponds to the schematic drawing in Fig. 8. Except for the input resistor R_i and input condenser C_i, all the elements in

the circuit were laid on the board as in Fig. 9. Since the control grid terminals of the 57 tubes are at the top, they were connected by a stiff wire from which the input condenser C_i was suspended and from which the grid leak R_i was joined to a conveniently placed ground wire. In order to be able to measure any plate current, pairs of binding posts were placed in all the plate leads of the three stages, as shown at X and Y. While it would have been convenient to have a milliammeter permanently in every plate circuit, it was felt to be an unnecessary expense. One can always connect a milliammeter to the binding posts in a particular circuit when it is desired to know the current in it. At other times the binding posts are connected by a short wire. Thus it is possible to make all the necessary observations with only three milliammeters. The input condenser of each stage, C_i, is provided with a Fahnestock binding post so that the stage may be readily isolated and studied by itself. Any resistor or condenser can be replaced easily with one of a different value, should one be interested in knowing the effect of such a change.
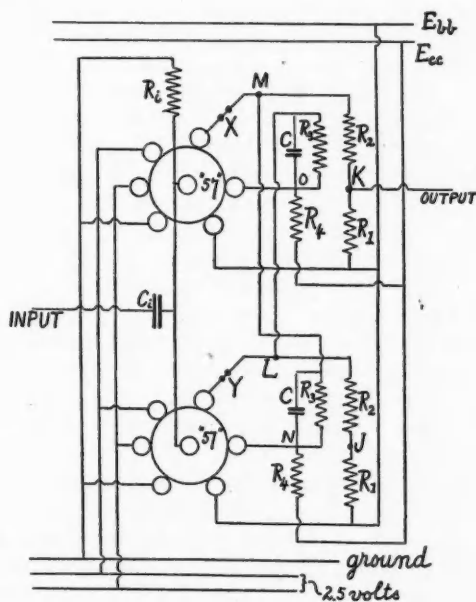

FIG. 9. One stage in the Reich scaling circuit, showing the locations of the various parts.

The values for the various elements used are as follows:

$E_{bb} = 45$ v $R_1 = 10,000$ ohms
$E_{cc} = -28$ v $R_2 = 50,000$ ohms
              $R_3 = 250,000$ ohms
$C = 50$ $\mu\mu$f $R_4 = 250,000$ ohms
$C_i = 50$ $\mu\mu$f $R_i = 100,000$ ohms

Except for the values of R_i and R_1, these are identical with those given by Reich. Condensers with $C_i = 50$ $\mu\mu$f work well between stages. This value, however, is too high for the input condenser between the control circuit and the first stage of the scale circuit. Here a radio "neutralizing" condenser provides sufficient capacitance (20 to 5 $\mu\mu$f).

The arrangement of the three stages on the board is shown in Fig. 10, where each "box" corresponds to Fig. 9. It was found convenient to use B-batteries to supply the plate and grid voltages. The batteries were located at the right-hand end of the board.

The voltage change of the point K, when the circuit trips, is about 6 v. This, while sufficient to trip a scaling circuit, is not enough to operate the Pickering recording circuit. The voltage change of the point M is more than sufficient but, unfortunately, a pulse of some sort comes back and upsets the scaling circuit. The remedy is to connect the last stage of the scaling circuit to the recording circuit through a simple amplifier as shown in Fig. 10. With a grid bias of 1.5 v (supplied by a miniature battery), the current through the 57 tube is 0.4 ma. This results in the normal potential of the point S having a value about 40 v below that of E_{bb}. A negative pulse coming from a third stage of the scaling circuit momentarily cuts off the current through the 57 tube, thereby allowing the potential of the point S to approach E_{bb}. This provides a positive pulse large enough to operate the recording circuit. (In most of the investigations, it is found better to connect M, rather than K, to the grid of the amplifier tube.)

3. *Characteristics of the scaling circuit.*—(a) Insight into the way the Reich scaling circuit works may be obtained by measuring the currents to the plates, and the voltages at the points J, K, L, M, N and O, Figs. 8 and 9. The former is accomplished by connecting milli-ammeters to the pairs of binding posts at X
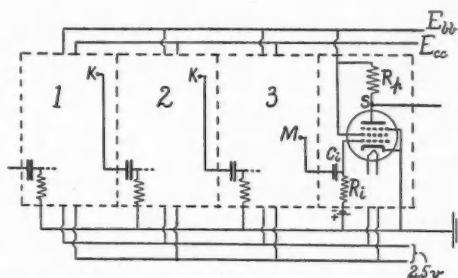

FIG. 10. Skeleton diagram of a 3-stage Reich scaling circuit with an amplifier to transmit the pulse from the third stage to the recording circuit. K and M correspond to the same symbols in Figs. 8 and 9. Over-all dimensions are 12×28 in. The input of the amplifier may be disconnected from the third stage and connected directly to the control circuit (Figs. 1 and 2) for scale-of-one investigations. $R_p = 100,000$ ohms; $R_i = 500,000$ ohms.

and Y. The milliammeters show clearly that when one tube is conducting the other is not, and that a suitable negative pulse will cause the conducting state to be transferred from one tube to the other. The voltages are measured by a Weston vacuum tube voltmeter[10] because an ordinary type voltmeter takes enough current to disturb the potential distribution. A typical potential distribution is given in Table I. As the conducting state is transferred from one tube to the other, the potential of K changes by about 6 v and this, when it gives a negative pulse, is sufficient to trip the next stage.

The voltage distributions may also be measured when the plate voltage E_{bb} is given other values, such as 22.5 and 90 v.

(b) It is most instructive to investigate the tolerances of the circuit, that is, the degree to which the various voltages can be changed without impairing the tripping of the circuit. The simple device for producing pulses of known voltages, as described in Section B,3(a) is connected to the input condenser and the range of values of the input pulse-voltage which will trip any one stage in the scaling circuit is investigated. Tripping is indicated by the reading

---

[10] The d.c. calibration curves supplied with the instrument were not sufficiently accurate. A new calibration was therefore made against a standard voltmeter and was found to be quite permanent. The instrument has 6 ranges giving nominal maximum voltages of 1.2, 3, 6, 8, 12 and 18 v. By means of a multiplier which takes a current of less than 10^{-6} amp, each range can be multiplied by 10, so that any voltage between 1 and 180 v can be measured with an accuracy of 1 percent.

of a milliammeter inserted at X or Y. With E_{bb} equal to 45 v, it is found that E_{cc} can be varied over a wide range without interfering with the power of a negative pulse to trip the circuit. Results illustrating this are tabulated in Table II. Similar investigations can be made when E_{bb} has other values, for example, 90 and 22.5 v.

(c) When the 3-stage scaling circuit feeds pulses into the recording circuit, only every eighth pulse is recorded. This means that if the mechanical recorder registers the arrival of n pulses, the number actually received by the scaling circuit may be anything between $8n$ and $8n+7$. The uncertainty may be removed by the following interpolation device. A milliammeter is inserted in the plate circuit of, say, the left-hand tube in each stage (Fig. 8). If we denote no current by "O" and a current by "X," it is easy to verify that the set of three milliammeters will pass through the configurations shown in Table III when the scaling circuit is operating properly. If, starting with an initial milliammeter configuration "O, O, O," the mechanical recorder registers, in a certain time interval, the arrival of 43 pulses with a final configuration of "X, O, X," it means that $(43 \times 8) + 6$ pulses had been received by the scaling circuit.

TABLE II. *Grid voltage tolerance.*

$E_{bb} = 45$ v	INPUT CONDENSER $= 50\mu\mu f$
E_{cc} (v)	INPUT PULSE FOR NORMAL TRIPPING (v)
-46	3 to >70
-30	2 to >70
-26	2 to >70
-22.5	1.5 to >70

TABLE III.

	STAGE 1	STAGE 2	STAGE 3
7	O	X	X
8	X	X	X
1	O	O	O
2	X	O	O
3	O	X	O
4	X	X	O
5	O	O	X
6	X	O	X
7	O	X	X
8	X	X	X
1	O	O	O
2	X	O	O

D. Experiments on counting without scaling

We now consider the observations that can be made with a G-M control circuit coupled directly to the recording circuit. This is sometimes known as a "scale of one." Since the recording circuit operates only on positive pulses, we connect its input condenser (B in Figs. 4 and 5) to the binding post in the control circuit which delivers positive pulses ($+B.P.$ in Figs. 1 and 2).

A way of using the *negative* pulses from the control circuit is to disconnect the input condenser of the amplifier tube from the scaling circuit (Fig. 10) and connect it to $-B.P.$ in Figs. 1 and 2. Thus the pulses pass from $-B.P.$ through the amplifier tube to the recording circuit. In some respects, particularly when comparing a scale of one with a scale of eight, this is more satisfactory than connecting the recording circuit directly to $+B.P.$

This combination affords an excellent opportunity for studying statistical fluctuations in the time intervals between the arrival of ionizing particles in the G-M tube. By tabulating the numbers of intervals occurring between 0 and 1 sec, 1 and 2 sec, etc., we get a distribution of time intervals given by experiment. This may then be compared with a Poisson distribution, which is the theoretical distribution derived on the assumption that the particles arrive at random. The derivation is given in Part II.

1. *Distribution of time intervals: comparison with the Poisson function* $p_0(t)$.—To get an experimental distribution it is advisable to place the radioactive source at such a distance from the G-M tube that the average time interval between pulses is about 3 or 4 sec. The simplest method of securing a set of data is to use a stop watch to measure the intervals between the pulses. Since it takes time to stop, reset, and start the stop watch, only alternate intervals between pulses can be measured. Care should be taken to record an interval, no matter how short it is, once the stop watch is started; this will counteract the tendency to disregard very short intervals. It is desirable to record at least 300 intervals, although it would be better to have a still larger number.

The chance that an interval with a length between t and $t+dt$ will occur is
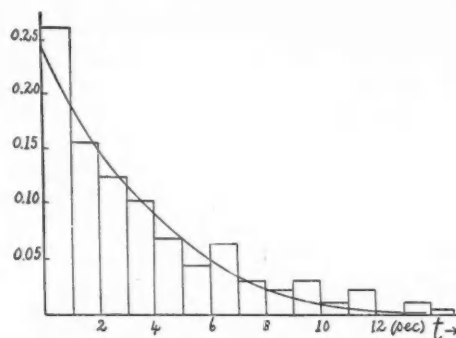
$$p_0(t)dt = (1/a) \exp (-t/a)dt,$$

FIG. 11. Distribution of random time intervals between pulses delivered by a G-M tube. Total number of intervals, 354; average interval a, 4.09 sec. The smooth curve gives the theoretical distribution, $p_0(t) = (1/a) \exp(-t/a)$. The rectangles give the experimental values of $p_0(t)\delta t$, where δt is 1 sec. For example, the intervals between 3 and 4 sec amount to 1/10 of all the intervals.

where a is the average time interval. Since a can be readily calculated, the curve $p_0(t)$ may be drawn. The next step is to count the intervals occurring in each of the ranges 0–1 sec, 1–2 sec, etc., and plot them, as fractions of the total number of intervals, on the $p_0(t)$ curve as shown in Fig. 11. This sort of representation is called a *histogram* and is particularly valuable when dealing with a finite number of observations characterized by statistical variations.

It is very instructive to take every other interval just as it comes in the data and so obtain two independent sets of time intervals. These can be treated separately and used to construct two histograms, which can be compared with each other and with Fig. 11. A comparison of this kind illustrates very clearly the fact that when we deal with a finite number of events, we must expect to find fluctuations in the observed number of events in each range. This is in no way to be attributed to experimental errors; it is inherent in the phenomenon itself. The larger the number of intervals we have to deal with, the more closely will the experimental values fit the theoretical curve. If this is not already evident from the comparison just suggested, it can be made so by comparing the histogram for all the data with that made from, say, one-fifth of the data. To diminish the effect of any possible progressive change in sensitivity, the latter can be obtained by using every fifth time interval.

It will be shown in Part II that if time intervals fit a Poisson distribution, the mean absolute departure from the mean should satisfy

$$|(t-a)|_{Av} = 0.736a.$$

If, then, the experimentally determined mean

departure $|(t-a)|_{Av}$ satisfies this equation, the set of values may belong to a Poisson distribution.[11] This test was applied to the set of intervals used for Fig. 11; $0.736a$ was found to be 3.01 sec, while the value of $|(t-a)|_{Av}$ gave 3.14 sec, an agreement which is about as good as one can expect with 354 intervals.

2. *Distribution of pulses per interval: comparison with the Poisson function $W_x(n)$.*—A more convenient way of examining the nature of the distribution of pulses is to count the number of pulses arriving in successive *equal* time intervals. This can be readily effected by placing a clock alongside the Cenco counter and recording the reading on the latter at equal intervals of, say, 20 sec. The number of pulses received in each interval is then found, from which it is possible to list the number of intervals during which 0, 1, 2, 3, ⋯ pulses arrive. Each number so obtained is divided by the total number of pulses, giving a fraction to be associated with 0, 1, 2, 3, ⋯ pulses per interval. These fractions are plotted in Fig. 12. These give us the proba-
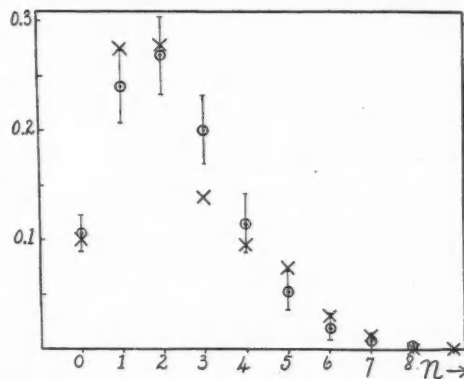

FIG. 12. Graph of $W_x(n)$, the distribution of intervals during which n (0, 1, 2, ⋯) pulses arrive, where x is the average number arriving per interval. Total number of pulses, 359; total number of intervals, 160; length of each interval, 15 sec. The circles are calculated from the theoretical formula for $W_x(n)$, and the vertical lines give the probable errors. The crosses are experimental values. No curve is drawn because $W_x(n)$ has no meaning except when n is an integer.

bilities of finding 0, 1, 2, ⋯ pulses in the interval.

According to the theory that is given in Part

[11] A more positive statement cannot be made. One could say, however, that if the relation is not satisfied the distribution of time intervals cannot be a Poisson distribution.

II, the chance of finding an interval during which exactly n pulses arrive is

$$W_x(n) = (1/n!)x^n \exp(-x),$$

where x is the average number of pulses arriving per interval, x being obtained by dividing the total number of pulses arriving by the number of intervals. In one experiment x was 2.24 pulses per interval. With this, the values of $W_x(n)$ were calculated for $n = 0, 1, 2, \cdots$. The probable error in each theoretical value is indicated by a short vertical line.[12] Four of our experimental points deviate from the theoretical value by less than the probable error, three by about the probable error and two by more than the probable error.

Another way to handle the observations from which Fig. 12 was constructed is as follows. In Part II it will be shown that if x is the average number of pulses arriving per interval and n is the number $(0, 1, 2, \cdots)$ actually arriving during successive intervals, the relation

$$(n-x)^2{}_{\mathrm{Av}} = x$$

must hold. Using the experimental values referred to, we find that $(n-x)^2{}_{\mathrm{Av}}$ is 2.60 while $x = 2.24$. Part, if not all, of the discrepancy can be attributed to the fact that the experimental data were obtained from 359 pulses distributed over 160 intervals. It is to be emphasized that statistical fluctuations make it impossible to secure exact agreement. However, the discrepancy will become less and less as the number of pulses and intervals in increased without limit.

3. *Losses in counting pulses: Volz-Schiff formula.*—The fact that a G-M counting unit made up of a control circuit and a recording circuit fails to record all the ionizing particles passing through the G-M tube, when their rate of arrival is not too small, may be shown in the following way. The number of counts recorded in a suitable time interval is plotted as a function of the reciprocal of the square of the distance of a

radioactive source from the G-M tube. If the unit counted every ionizing particle, then, apart from statistical fluctuations, the curve obtained would be a straight line passing through the origin. It is found experimentally, however, that at first the curve is tangent to a straight line through the origin, then passes through a maximum, after which it decreases to zero, indicating that more and more counts are lost as their number increases until the counting unit is completely "jammed" or "paralyzed." While no part of the unit has an instantaneous response, the reaction time of the mechanical counter is usually sufficiently longer than that of any other part of the circuit for the counter to be considered as responsible for the losses. Some mechanical counters, such as the Cenco counter, are so constructed that they require a definite interval of time, during which no pulse arrives, in order to reset themselves for another pulse. Let k be this time interval. Volz[13] and Schiff[14] have shown that the relation between N, the number of recorded counts, and N_0, the number of pulses received by the mechanical counter, is $N = N_0 \exp(-N_0 k)$. At first N increases with N_0, then passes through a maximum after which it decreases because of the exponential factor. While the implications of this equation could be tested, as indicated earlier in this paragraph, by using the inverse square law to control the number of ionizing particles passing into the G-M tube, it is more satisfactory to use a scaling circuit to determine N_0. The method is outlined in the next section.

E. Experiments on counting with scaling

To demonstrate the action of the scaling circuit in our unit two methods may be followed. (a) The positive pulse binding post in the control circuit ($+B.P.$ in Fig. 1) is connected directly to the input of the recording circuit while the negative pulse binding post ($-B.P.$ in Fig. 1) is connected to the input of the scaling circuit. Milliammeters are inserted in the plate circuits of the corresponding tubes in each of the three stages of the scaling circuit. With this arrangement it may be shown that, for every pulse registered by the recording circuit, the first

---

[12] Any one observation deviates from the mean of a very large number of observations by its *error*. The chance that this error is as likely as not to be larger than a certain quantity defines this quantity as the *probable error*. The *relative standard error* of a value obtained from a finite number of counts can be shown to be equal to the reciprocal of the square root of the number of counts. The probable error is obtained by multiplying this by 0.67. If, for example, one ordinate in Fig. 12 corresponds to an average of 100 counts per selected interval, the standard error is 10 percent and the probable error is 6.7 percent.

[13] H. Volz, Zeits. f. Physik 93, 539 (1935).

[14] L. I. Schiff, Phys. Rev. 50, 88 (1936).

stage trips once, the second stage half as often, and the third stage once every four times. (b) Alternatively, $-B.P.$ of Fig. 1 may be connected to the input of the amplifier stage, shown to the right in Fig. 10, which in turn is connected to the recording circuit. In both (a) and (b) the third stage of the scaling unit is disconnected from the amplifier stage.[15]

1. *Adjustment.*—The control circuit and scaling circuit are placed close to each other along the edge of a table with the recording circuit a few inches behind the scaling circuit. Certain adjustments are necessary to secure normal performance. The coupling condensers at the input to each stage of the scaling circuit are 50 $\mu\mu$f each. These behave satisfactorily with the artificial pulse device previously described. However, when connected to the $(-B.P.)$ output of the control circuit, some or all of the stages appear to be paralyzed. This appears to be due to the input pulse being too large. One remedy is to diminish the capacitance of the input condenser to the first stage to a value between 20 and 5 $\mu\mu$f. This is easily done by means of a variable radio "neutralizing" condenser. Another remedy is to insert a 100,000-ohm resistor between the megohm resistor (Y, Fig. 1) and ground, and take off the smaller pulse appearing at the junction of the two resistors.

The method of testing whether or not normal scaling occurs is to feed pulses from the control circuit simultaneously to the scaling and recording circuits. The recording circuit may be connected as in (a) or (b) above; (b) seems to be preferable. It may happen that scaling will be normal; that is, the first stage, as indicated by its milliammeter, will trip every time the recording circuit registers a pulse and the other stages will trip every other and every fourth time, respectively. (By "tripping" here is meant a change of a milliammeter reading in either direction; the succeeding stage is operated every alternate time one stage trips.) Should the scaling be irregular, it may be necessary to shield the scaling circuit by the grounded metal box referred to in Section **C**,2. Even though

[15] If *two* recording units are available, one could follow the method of Lifschutz and Duffendack and connect the output of the scaling circuit to one unit and the output of the control circuit to the other. This is a better experimental arrangement, but it ties up more equipment.

this box does not provide complete shielding, because a long narrow opening must be left on the front side to allow connections to be made to the three milliammeters just in front, it is quite adequate.

It is instructive, however, to observe how this shielding may become inadequate. On one occasion it happened quite unintentionally that a short wire connected to a sheet of insulated metal was left connected to the input of the recording circuit. This acted as a small aerial and disturbed the action of the scaling circuit in spite of the shielding which was adequate under normal working conditions. The conclusion to be drawn is that when a scaling circuit does not work properly it may be that the shielding is inadequate.

On placing a radioactive source at some distance from the G-M tube so that the Cenco counter is recording quite slowly, it is possible to verify the relation between the frequency of tripping in the first, second and third stages of the scaling circuit and the frequency of recording in the Cenco counter. At the same time the method of using the indications of the milliammeters to provide interpolated readings may be compared with the scheme given in Table III.

2. *Losses in counting with and without scaling.* —Formulas and curves, showing how scaling reduces counting losses, are given in Part II. A method of testing some of them is as follows. The relative strength of the radiation falling on a G-M tube from a radioactive source may be determined by means of the inverse square law, as long as the distance between the source and the tube is sufficiently greater than the dimensions of the latter. The scaling circuit, with its three stages operating normally, is connected between the control and recording circuits. The number of pulses recorded in a suitable time interval is determined for various positions of the radioactive source with respect to the G-M tube. As long as the recorded number is proportional to the intensity of radiation as determined by the inverse square law, we may assume that no counts are lost. On multiplying the recorded number by 8 (the scaling factor), we obtain the number of pulses fed into the scaling circuit by the control circuit for each position of the radioactive source. The scaling circuit is now removed and the control circuit connected directly to the recording circuit or through the amplifier as outlined at the beginning of Section
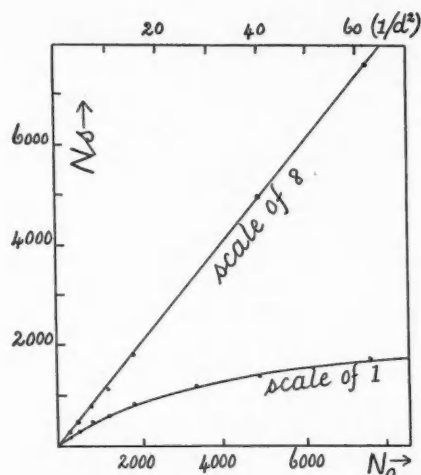
FIG. 13. Comparison of Ns, the "inferred number of counts" obtained with a scale of one and a scale of eight. Range of N_0, the number of pulses delivered by the control circuit, is from 0 to about 8000 per min. As checked by the inverse square law (see top of diagram), no losses occur with a scale of eight, while there are considerable losses with a scale of one.

E. The number of counts is again measured with the radioactive source in each of the positions mentioned. The results are shown in Figs. 13 and 14. Here the ordinates are the values of Ns, the product of the number of recorded counts and the scaling factor s, which is 8 with the scaling circuit in action and 1 without it. The abscissas are proportional to the inverse square of the distance between the radioactive source and the G-M tube. The straight portion of the curve for the scale of eight allows us to label the abscissas also in terms of $N_0(=Ns$, with $s=8)$, the number of pulses furnished by the control circuit. The inverse square law makes it possible for us to extrapolate N_0 beyond the range over which the scale of eight gives a straight line.

If the Volz-Schiff theory for the scale of one applies to our experimental results, it should be possible to fit our curve to the formula

$$N = N_0 \exp(-N_0 k).$$

By comparing our experimental curve (labeled $s=1$) in Fig. 13 with the theoretical curve in Fig. 20, it is seen that they agree qualitatively, at least so long as the scale-of-one curve is rising. Quantitatively the agreement is not good, for the values of k, instead of being constant, vary between 0.015 and 0.010 sec for different parts of the scale-of-one curve. The lack of constancy will be discussed later. Meanwhile it is interesting to consider the implications of a value of $k=0.015$ sec. On inserting this into the formula we find that for N_0 equal to 500 and 200 pulse/min, the corresponding values of N are 441 and 190, respectively. This means that, for counting rates even as low as 500 and 200 pulses/min, the percentage losses are as much as 12 and 5, respectively. This is surprising when one considers that the mean time intervals between pulses are as much as 0.12 and 0.30 sec for the two cases considered. As these times are much larger than the reaction time, $k=0.015$ sec, one is inclined to think that the counter would not miss any pulses. The explanation is that, with a random distribution of pulses, the shorter time intervals are relatively the more numerous, and it is the shorter intervals that are responsible for the losses. As Lifschutz and Duffendack have pointed out, an important advantage of a scaling circuit is that with it the number of short intervals leading to losses is greatly reduced.

The lack of quantitative agreement between the experimental curves in Figs. 13 and 14 and the theoretical curves in Figs. 20 and 21 calls for some comment, but it is not within the scope of this paper to attempt a detailed analysis of the discrepancy. As we have stated, the scale-of-one curve cannot be fitted closely to the equation $N=N_0 \exp(-N_0k)$. Instead of being constant, k varies from about 0.015 to about 0.010 sec as we go along the curve from the origin to near the maximum. It is likely that the experimental conditions are more complicated than those contemplated in a simple theory. In Lifschutz and Duffendack's recording circuit, the reaction time of the electric circuit was much smaller than that of the Cenco counter and so all the losses in the recording circuit may be assumed to be in the counter. In the Pickering circuit used in this experiment the RC value of the electric circuit turns out to be 0.005 sec, which is smaller than the reaction time of the Cenco counter but by no means of a different order of magnitude. Ruark and Brammer[16] have shown that there are two different ways in which losses can occur. In Type I recorders, to which the Cenco counter belongs, a minimum time k must elapse between pulses before the recorder can reset itself and be ready to record a second pulse. If the pulses come in very rapidly the recorder is paralyzed because no time interval is as long as k, and

[16] A. E. Ruark and F. E. Brammer, Phys. Rev. 52, 322 (1937).

the recorder never has a chance to reset itself. In Type II recorders, a pulse starts a cycle which is completed irrespective of whether or not other pulses arrive during the cycle. Such recorders do not paralyze, but record at a constant maximum rate when the number of pulses fed into them exceeds a certain value. The Pickering recorder is a Type II recorder. Our recording unit is, therefore, a combination of Type I and Type II devices with reaction times that are not widely different. This probably is the reason that the Volz-Schiff formula does not apply quantitatively to our scale-of-one results. A change to the type of recording circuit used by Lifschutz and Duffendack, in which the reaction time of the electric circuit is so fast that the Cenco counter alone determines the losses, would no doubt permit a quantitative check of the Volz-Schiff formula. However, the simple Pickering circuit is quite adequate for the other experiments described in this paper.

An instructive experiment is to reconnect the scaling circuit so that one can go through the same steps using a scale of two and a scale of four. The results may be plotted as in Figs. 13 and 14 along with the results for the scales of one and eight. The progressive decrease in the counting losses as the scaling factor is increased is made evident by such curves.

3. *Distribution of time intervals recorded by a scaling circuit.*—It has been pointed out that a scaling circuit not only increases the average time between the recorded pulses but also reduces the number of short intervals very greatly, which is equally if not more important.[17] It is the presence of short intervals that leads to counting losses.

It is instructive to show this by direct experiment. The scale-of-eight circuit is introduced between the control and the recording circuits. The radioactive source is placed at such a distance that the average time between recorded counts is about 5 sec. Then, with a stop watch, several hundred time intervals are measured. As in Section **D**,1, because of the time necessary to stop, reset and start the stop watch, the observer should record every alternate interval. These intervals can then be classified and the number in the ranges 0–1, 1–2, $\cdots$ sec plotted as a histogram in Fig. 15. The theoretical curve is given by

$$y = p_7(t)dt = (1/a)(t/a)^7(1/7!) \exp(-t/a)dt,$$

which measures the chance that the eighth pulse received by the scaling circuit will come within an interval between t and $t+dt$. (The

[17] H. Lifschutz and O. S. Duffendack, Phys. Rev. **54**, 714 (1938); L. Alaoglu and N. N. Smith, Phys. Rev. **53**, 832 (1928).
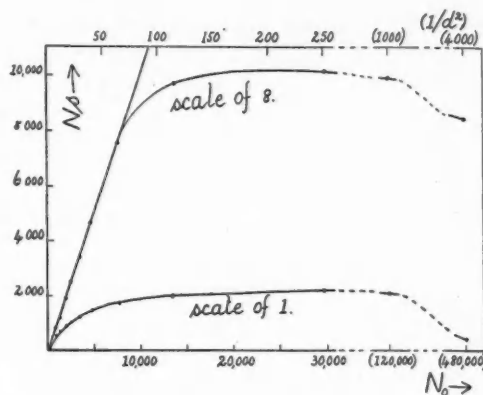
FIG. 14. Comparison of Ns, the "inferred number of counts," obtained with a scale of one and a scale of eight when N_0, the number of pulses delivered by the control circuit, ranges from 0 to presumably over 100,000 per min.

general formula for y is Eq. (21) in Part II; here we have substituted 7 for n and $1/a$ for λ.) This curve is plotted in Fig. 15.

It is not difficult to see that the larger the scaling factor, the more closely grouped around their average value are the time intervals between the pulses delivered by the scaling circuit. In particular, as the scaling factor increases, the proportion of short intervals—those responsible for losses in counting—very rapidly becomes smaller. With a sufficiently large scaling factor, no counts will be lost in the recording circuit. Losses may occur, however, in the G-M tube itself or in its control circuit.

II. STATISTICS OF COUNTERS AND ASSOCIATED CIRCUITS

A brief account of that part of statistics bearing on counters and their associated circuits will now be given. When particles arrive in a counter at random, the distribution of time intervals between successive arrivals is called a *Poisson distribution* or occasionally, a *Bateman distribution*. If the particles are ionizing agents, a perfect counter will deliver an electric pulse into a recording device, or into a scaling circuit, for each particle arriving. If all parts of the system responded positively and instantaneously we should obtain a record of, or a "count" for, each ionizing particle. However, no counting system is able to record faithfully the arrival of

successive particles as the speed of arrival is increased without limit. This is because the counting system requires a small but finite time to reset itself after it has responded to a particle. The number of recorded "counts" will be less than the number of particles arriving when the time interval between the arrival of successive particles does not always exceed a certain value. The failure to record all the particles arriving usually may be attributed to the finite time required by the mechanical counter to reset itself. Thus the Cenco impulse counter will not record two counts when two pulses arrive within about 0.008 sec of each other. With a suitable scaling circuit, however, it is possible to obtain conditions under which the slowness of response of the impulse counter is no longer a factor in determining the number of counts missed. The number missed will now be determined by the time constants of the electric circuits or of the Geiger-Müller tube itself.

Since the ionizing particles arrive at random, a quantitative discussion of the counts lost under various conditions clearly involves statistical considerations. The first thing to do is to set up the distribution formulas $W_x(n)$ and $W_n(t)$: $W_x(n)$ is the probability of arrival of n particles within an arbitrarily selected time interval, where x is the *average* number arriving in the interval; $W_n(t)$ is the probability of the arrival of n particles in any time t. In $W_x(n)$ attention is focused on the number arriving during a selected interval, whereas in $W_n(t)$ attention is focused on the way the chance of arrival of n particles varies with the interval t.

A. Derivation of $W_n(t)$ and $W_x(n)$

If λ be the average number of particles arriving per second, λ is clearly $1/a$, where a is the average time between the arrival of successive particles. The chance of a particle arriving within a time element dt is λdt. The chance of a particle *not* arriving within element dt is $(1-\lambda dt)$. Let $W_n(t)$ be the chance of n particles arriving in an interval t, and let $W_{n+1}(t+dt)$, be the chance of $n+1$ particles arriving in an interval $t+dt$. $W_{n+1}(t+dt)$ is made of two chances, which are (a) the chance of $n+1$ particles arriving in time t followed by none arriving in element dt, and (b) the chance of n particles arriving in time t followed by one particle in element dt. Chance (a) may be written as $W_{n+1}(t) \times (1-\lambda dt)$ and chance (b) as $W_n(t) \times \lambda dt$. We then have

$$W_{n+1}(t+dt) = (1-\lambda dt)W_{n+1}(t) + \lambda dt W_n(t). \quad (1)$$

By Taylor's theorem

$$W_{n+1}(t+dt) = W_{n+1}(t) + dt(d/dt)W_{n+1}(t). \quad (2)$$

On putting Eq. (2) into (1) we get

$$(d/dt)W_{n+1}(t) = \lambda[W_n(t) - W_{n+1}(t)]. \quad (3)$$

To solve Eq. (3) we put $n+1 = 0, 1, 2, \cdots$, and get

$$(d/dt)W_0(t) = -\lambda[0 - W_0(t)], \text{ since } W_{-1}(t) = 0,$$
$$(d/dt)W_1(t) = -\lambda[W_0(t) - W_1(t)],$$
$$(d/dt)W_2(t) = -\lambda[W_1(t) - W_2(t)].$$

The solutions are

$$W_0(t) = \exp(-\lambda t),$$
$$W_1(t) = (1/1!)(\lambda t)\exp(-\lambda t),$$
$$W_2(t) = (1/2!)(\lambda t)^2 \exp(-\lambda t)$$

and, in general,

$$W_n(t) = (1/n!)(\lambda t)^n \exp(-\lambda t). \quad (4)$$

To recapitulate,

$W_0(t)$ is the chance of 0 particles arriving within an interval t,

$W_1(t)$ is the chance of 1 particle arriving within an interval t,

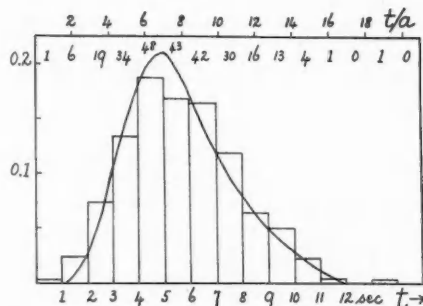$W_n(t)$ is the chance of n particles arriving within an interval t.

Fig. 15. Distribution of time intervals delivered by the scale of eight when fed by random pulses from the control circuit. The smooth curve gives the theoretical distribution, $p_7(t) = (1/a)(t/a)^7(1/7!)\exp(-t/a)$, where a, the average time interval between pulses *received by* the scaling circuit, is 0.718 sec. The rectangles give the experimental values of $p_7(t)\delta t$, where δt is 1 sec. For example, the intervals between the pulses *delivered by* the scaling circuit, having values between 4 and 5 sec, amount to the fraction 0.185 of *all* the intervals. Above each rectangle is indicated the number of experimental observations contributing to it.

In the foregoing discussion, $W_n(t)$, a function of the time, tells us the chance of finding an interval of time t during which n particles arrive. We now change our point of view and ask what is the chance that 0, 1, 2, $\cdots$ n particles will arrive within some arbitrarily chosen time interval t (now held constant). Let x be the *average* number of particles arriving in the interval t. Then, from the definition of λ, we have $x = \lambda t$. Eq. (4) becomes.

$$W_x(n) = (1/n!)x^n \exp(-x). \qquad (5)$$

The foregoing derivations are based on those given by Bateman.[18]

B. A special case: $W_0(t)$

The chance of finding a time interval t during which no particle arrives is given by

$$W_0(t) = \exp(-\lambda t). \qquad (6)$$

It tells us that the most probable value of a time interval during which no particle arrives is zero, a result which is sometimes considered surprising.

It will be recalled that Eq. (6) is the equation we meet in the kinetic theory of gases for the probability of free paths. It can be equally well thought of as giving the chance of *no collisions* occurring with a path of *length* t, when the average number of collisions per unit distance is λ. Since, by definition, such a path is a free path, we see that Eq. (6) describes the chance of occurrence of a free path at least equal to t in length, when the mean free path is $1/\lambda$.

C. $W_n(t)$: Graphs and maximum values

The curves for $W_n(t)$ for values of n from 0 to 4 are shown in Fig. 16. There the various func-
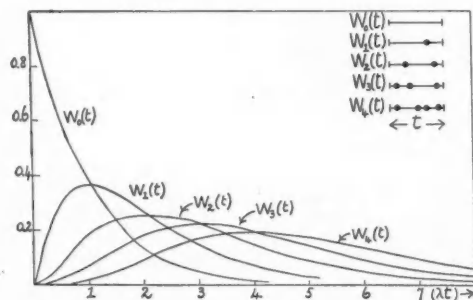

Fig. 16. Graphs of $W_n(t)$, for $n = 0, 1, 2, 3$ and 4.

[18] H. Bateman, Phil. Mag. **20**, 704 (1910); **21**, 745 (1911).

tions are plotted to the same scale, while in Fig. 17 we show how the curve for $W_{15}(t)$, when suitably adjusted so that its maximum coincides with that of $W_3(t)$, is *relatively* narrower than that for $W_3(t)$. [$W_{15}(t)$ and $W_3(t)$ were selected because they are involved in the characteristics of a scale-of-sixteen circuit and a scale-of-four circuit, respectively.] Tedious calculations may be avoided since these functions are tabulated.[19, 20] By differentiating Eq. (4) with respect to t, and equating to zero, we find values for the "most probable" time intervals t_m to be given by

$$n - \lambda t_m = 0. \qquad (7)$$

This tells us that the most probable values of the time intervals during which just 0, 1, 2, $\cdots$ particles arrive are 0, $1/\lambda$, $2/\lambda$, $\cdots$, respectively.
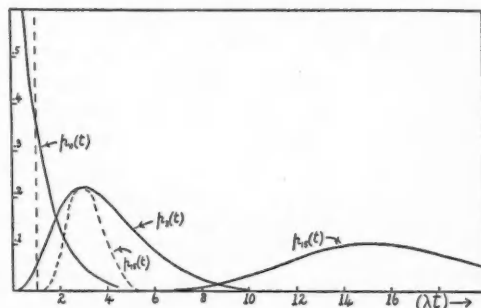

FIG. 17. Graphs of $p_n(t)$ for $n = 0, 3$ and 15, corresponding to scales of one, four and sixteen, respectively. The full lines are all to the same scale. The dotted curve $p_{15}(t)$ is adjusted to match the curve $p_3(t)$ at its maximum. A curve $p_n(t)$ is identical in *shape* with a curve $W_n(t)$, the only difference being that the ordinates of the former are λ times those of the latter, where λ is the number of pulses per second.

D. The nature of $W_x(n)$

In the preceding discussion, $W_n(t)$ is a function of the time which tells us the chance of finding an interval of time during which exactly n particles arrive (no more and no less); n is a parameter which is held constant while t is the variable. We now look at the situation somewhat differently and ask what is $W_x(n)$, the chance that any selected number of particles—for example, 0, 1, 2, $\cdots$ n— will arrive within an arbitrary time interval t, which is held constant. Here n is the variable, which, however, can assume only

[19] K. Pearson, *Tables for Statisticians and Biometricians*, (Cambridge Univ. Press), Part I, pp. 113–21.
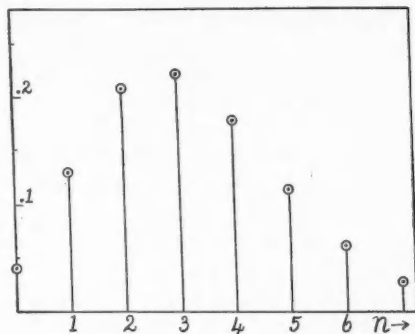[20] T. C. Fry, *Probability* (Van Nostrand), pp. 458–463.

FIG. 18. Graphs of $W_x(n)$, the chance of finding an interval of time during which exactly n pulses arrive, where x is the *average* number of pulses per interval. The diagram is drawn for $x=3.20$ pulse/interval.

positive integral values. The form of $W_x(n)$ is given in Eq. (5). In order to plot its values for various n's, it is necessary to assume some value for x, such as 3.2 particle/sec. This is done in Fig. 18, which, among other things, shows that the chance of arrival of any number of particles between 0 and 6 per sec amounts to 95.5 percent, leaving 4.5 percent for the chance of arrival of 7, 8, $\cdots$ particle/sec.

We cannot expect exact agreement with the results shown in Fig. 18, unless we make an enormous number of observations. If, for example, the ordinate for $n=2$ represents the experimental value from an exceedingly large value of observations, then the probable error in the value of the ordinate deduced from 10,000 of these would amount to $0.67/(10,000)^{\frac{1}{2}}$ or about 0.67 percent.

E. $W_x(n)$: properties, mean square deviation

(*a*) The chance that any number of particles from $n=0$ to $n=\infty$ will arrive during a selected time interval must be unity. That $W_x(n)$ satisfies this may be seen by adding up the chances that 0, 1, 2, $\cdots$ arrive. This follows from

$$\sum_{n=0}^{\infty} x^n \exp(-x)/n!$$
$$= \exp(-x)(1+x/1!+x^2/2!+\cdots)$$
$$= \exp(-x)\exp(+x)=1.$$

(*b*) The *average* number of particles arriving during the selected interval is obtained by multiplying n, the number arriving during any one of these intervals, by $W_x(n)$, the chance of arrival of exactly n particles during the interval,

and then summing from $n=0$ to $n=\infty$. Hence

$$\bar{n}=\sum_{n=0}^{\infty} nx^n \exp(-x)/n!$$
$$=\sum_{n=0}^{\infty} \exp(-x)[x^n/(n-1)!]$$
$$= \exp(-x)(1/\infty+x/0!+x^2/1!+x^3/2!+\cdots)$$
$$= \exp(-x)\times x\times \exp(+x)=x.$$

We thus verify that x is the average number arriving per interval. Of course, x was defined as the average number; what we have done is to show that the formula is consistent.

(*c*) We shall show that the mean-square deviation from the mean, of the particles arriving during any arbitrarily selected interval, is equal to the mean. We have to show that

$$(n-x)_{Av}^2=x. \tag{8}$$

The chance of n particles arriving during the interval is $W_x(n)$. The corresponding deviation from the mean is $(n-x)$ and its square is $(n-x)^2$. The product $(n-x)^2 W_x(n)$ is the chance that this particular deviation, squared, appears. The *mean* squared deviation is obtained by summing this quantity from $n=0$ to $n=\infty$, or

$$(n-x)_{Av}^2=\sum_{n=0}^{\infty}(n-x)^2 x^n \exp(-x)/n!$$
$$=\sum_{n=0}^{\infty}(n^2-2nx+x^2)x^n \exp(-x)/n!$$
$$= \exp(-x)\left[\sum \frac{n(n-1)+n}{n!}x^n\right.$$
$$\left.-2x\sum \frac{n}{n!}x^n+x^2\sum \frac{x^n}{n!}\right]$$
$$= \exp(-x)\left[\sum \frac{x^n}{(n-2)!}+\sum \frac{x^n}{(n-1)!}\right.$$
$$\left.-2x\sum \frac{x^n}{(n-1)!}+x^2\sum \frac{x^n}{n!}\right]$$
$$= \exp(-x)[x^2 \exp(+x)+x \exp(+x)$$
$$-2x^2 \exp(+x)+x^2 \exp(+x)]=x.$$

Therefore we have derived a very important result, namely, Eq. (8), which is characteristic of the Poisson distribution. This asserts that, if we

count the number of particles arriving in any arbitrarily selected time interval, compute the mean number arriving per interval, tabulate the deviations from the mean and square then, the average of the squares of the deviations from the mean turns out to be the mean itself. This is typical of the Poisson distribution and can be used to test whether or not an observed distribution is a Poisson distribution. If, for example, the distribution of *time* intervals between successive particles were not the Poisson distribution $W_0(t)$, but something like $W_5(t)$, then the mean value of $(n-x)^2$ would certainly *not* be x.

F. $W_0(t)$: mean deviation

Suppose that we record the time intervals between successive particles coming in at random. The probability of such time intervals is clearly $W_0(t)$, since they are intervals during which no particle arrives. The mean interval is denoted by a, which is the reciprocal of λ. We can show that the absolute deviation from the mean $|(t-a)|_{Av}$ can be expressed very simply.

Let $U_0(t)$ be the chance of a time interval equal to or larger than t, during which no particle arrives. $U_0(t+dt)$ is equal to $U_0(t)$ multiplied by the chance of no particle arriving in element dt, which is $(1-\lambda dt)$. Hence

$$U_0(t+dt) = (1-\lambda dt) U_0(t),$$

from which it follows, much as in Section **A**, Part II, that

$$U_0(t) = \exp(-\lambda t). \qquad (9)$$

It can be argued that $U_0(t)$, the chance of occurrence of an interval equal to, or larger than, t, characterized by the arrival of no particles is identical with $W_0(t)$, the chance of arrival of no particles during a time interval t. The derivation leading to Eq. (9) is given as an alternative to this argument.

The chance of there being an interval between t and $t+dt$ during which no particle arrives is proportional to dt and may be written as $p_0(t)dt$. It is clear that

$$p_0(t)dt = U_0(t) - U_0(t+dt) = dt(d/dt) U_0(t),$$

from which

$$p_0(t) = \lambda \exp(-\lambda t) = (1/a) \exp(-t/a), \quad (10)$$

since the average time interval between the arrival of particles is equal to the reciprocal of the average number of particles arriving per unit time. The deviation of any *one* interval t from the average a is $a-t$ if $t<a$, and $t-a$ if $t>a$. To get the "mean absolute deviation from the mean" we multiply each value of $a-t$ or $t-a$ by the chance of occurrence of that interval, which is $p_0(t)dt$, and integrate from 0 to ∞:

$$|(t-a)|_{Av} = \int_0^a (a-t)(1/a) \exp(-t/a)dt$$

$$+ \int_0^a (t-a)(1/a) \exp(-t/a)dt$$

$$= 2a/e = 0.736a. \qquad (11)$$

This can be used as a criterion to test whether or not a large number of time intervals belong to a Poisson distribution. If Eq. (11) is satisfied, then the set of intervals may belong to a Poisson distribution; but if it is not satisfied, then the set of intervals cannot belong to a Poisson distribution. It is to be understood that, even with a Poisson distribution, exact agreement with Eq. (11) can be expected only when the set of observations contains a very large number of values.

G. Scaling: chances of double, triple, . . . intervals

A scaling circuit is a device to feed, into a suitable mechanical recorder such as a Cenco impulse counter, every 2nd, 3rd, ··· pulse, depending on the "scale." A "scale of one" feeds in every pulse, a "scale of s" skips n pulses and feeds in the $(n+1)th$, so that $s=n+1$. We have seen that if particles come in at random and if each particle generates a pulse, the chance of an interval t containing

0 pulse is $W_0(t) = \exp(-\lambda t)$,
1 pulse is $W_1(t) = (1/1!)(\lambda t) \exp(-\lambda t)$,
2 pulses is $W_2(t) = (1/2!)(\lambda t)^2 \exp(-\lambda t)$,
n pulses is $W_n(t) = (1/n!)(\lambda t)^n \exp(-\lambda t)$.

Curves for these functions are shown in Fig. 16.

We now introduce a new function, $y_n(t)$, which measures the chance of an interval t containing any number of pulses from 0 to n, but not over n. It is clear that
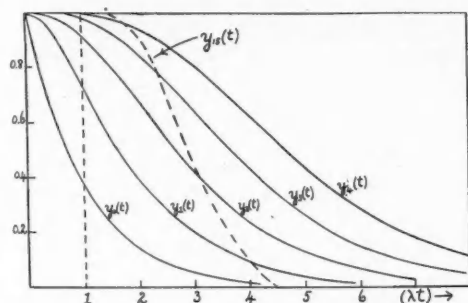
FIG. 19. Graphs of $y_n(t)$, the chance of an interval t containing any number of pulses from 0 to n. The curve $y_{15}(t)$ is dotted to indicate that the scale of its abscissas has been reduced fivefold.

$$y_0(t) = W_0(t),$$
$$y_1(t) = W_0(t) + W_1(t),$$
$$y_n(t) = W_0(t) + W_1(t) + \cdots + W_n(t), \quad (12)$$

$$y_n(t) = \exp(-\lambda t)$$
$$\times [1 + (\lambda t)/1! + (\lambda t)^2/2! + \cdots (\lambda t)^n/n!]. \quad (13)$$

Fry[19] tabulates values of a certain function, which, on being subtracted from unity, give numerical values for $y_n(t)$. This makes it unnecessary to go through some rather tedious arithmetic to compute $y_n(t)$.

The values of $y_n(t)$ for various n's are plotted in Fig. 19. It will be observed that, as n increases, the values of $y_n(t)$ tend to keep closer to their maximum values for larger and larger values of t. It is this peculiarity which is significant in scaling circuits, for we shall see that by means of it the number of counts lost can be greatly diminished by the use of a suitable scaling circuit.

Let us draw an ordinate of unit length at $\lambda t = \lambda k$, where k is an arbitrarily selected time interval. (In Fig. 19 we have shown this ordinate as a broken line at $\lambda t = \lambda k = 1$, in order to give numerical illustrations.) Each of the curves $y_n(t)$ cuts this vertical line in two parts. The lower part, from the abscissa up to $y_n(k)$, measures the chance that any number of pulses from 0 to n, but not over n, will arrive within $t = k$. The upper part, from $y_n(k)$ to the horizontal line, measures the chance that $(n+1), \cdots$ to ∞ pulses will arrive within $t = k$. It is clear that, as n increases, the chance that any number of pulses up to n will arrive during $t = k$ increases very rapidly. In the specific case illustrated in

Fig. 19 the chance that no pulse will arrive during the interval is 0.37, while the chance that from 0 to 4 pulses, inclusive, will arrive is 0.99. Alternatively, this may be expressed by saying that the chance of an interval being characterized by the arrival of *more* pulses than n decreases rapidly as n increases. This is more obviously true for the left portion of Fig. 19, but this is the portion of most significance in our application.

H. Counting losses: scale of one, Volz-Schiff formula

Since all the elements in a counting system, including the G-M tube, its control circuit, the recording circuit and the mechanical impulse counter into which it feeds pulses, have various finite times of response, it is clear that, as the number of particles arriving in the G-M tube increases, the number actually counted will differ more and more from the number arriving. It should be pointed out that the pulses can be "lost"—that is, not counted by the impulse counter—in at least two ways.[21] (1) It may happen that when an element in the counting system has received a pulse, it goes through a cycle occupying a definite time $t = k$ during which it is insensitive to any pulses that may arrive, before it is ready to record a second pulse; if the number of pulses arriving per second be increased without limit, such an element will count merely at its maximum rate of $1/k$ per sec, a rate that it cannot exceed. (2) It may happen that an element requires a time interval $t = k$, during which *no* pulse arrives, before it can reset itself to register a pulse. Any pulse that arrives within $t = k$ of the time at which the preceding pulse arrived, defers the time by $t = k$ at which a succeeding pulse can be recorded. Thus, if the pulses follow each other too quickly, so that the intervals are less than k, this element never gets a chance to reset itself and so will not count at all. It is said to be "paralyzed." The Cenco impulse counter is of the later kind. Some electrical elements are of the former kind. As a rule, the mechanical impulse counter has a much longer time of response than any other element in the counting system. In the absence of a scaling

[21] A. E. Ruark and F. E. Brammer, Phys Rev. **52**, 322 (1937); H. Lifschutz and O. S. Duffendack, Phys. Rev. **54**, 714 (1938).

circuit, it is therefore usually justifiable to consider the counter as the part of the system in which the counting losses occur.

Let N_0 be the *average* number of pulses delivered to the impulse counter per second. If there were no losses, the number of counts recorded, N, would be equal to N_0. But whenever a time interval between pulses less than $t=k$ occurs, the counter does not record them as two counts. In fact, only when an interval between pulses equals or exceeds k is each pulse recorded. Pulses separated from preceding pulses by less than k are not recorded. The number of pulses counted may readily be obtained from Fig. 19. It will be remembered that, in the discussion of these curves in Section **G**, the length of the ordinate at $\lambda t = \lambda k$ from the abscissa up to the curve $y_0(t)$ measures the chance of an interval between pulses being at least equal to k.[22] The value of this ordinate thus gives us the fraction of the total number of pulses arriving that are recorded. We then have

$$N = N_0 y_0(k) = N_0 \exp(-\lambda k) \qquad (14)$$

and, *since N_0 is identical with λ*, we can write N as

$$N = N_0 \exp(-N_0 k). \qquad (15)$$

This equation was derived independently by Volz and by Schiff. The resulting curve is designated by $s=1$ in Figs. 20 and 21. The maxi-
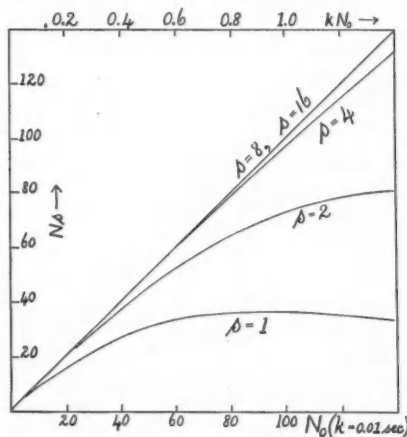

FIG. 20. Counting losses for various scaling factors. *Ordinates* are values of Ns, the "inferred number of counts." *Abscissas* are the true number of counts N_0 for a counter with a reaction time k of 0.01 sec.

[22] See also Section **B**.
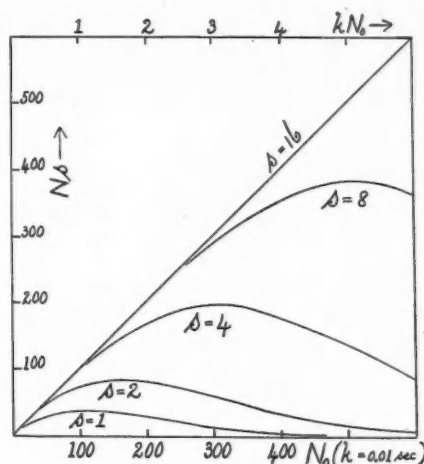
FIG. 21. Counting losses as in Fig. 20, but with a fivefold larger range in N_0.

mum ordinate $N_{\max}$ is determined by

$$k = 1/eN_{\max} \qquad (16)$$

which, on substitution in Eq. (15), yields

$$N = N_0 \exp(-N_0/eN_{\max}). \qquad (17)$$

As N_0, the number of pulses fed into the impulse counter, increases, N, the number recorded, increases to a maximum and then falls off. If we can determine $N_{\max}$ experimentally (for example, by moving a radioactive source in the vicinity of the G-M tube until the impulse counter is recording at its maximum rate, which is $N_{\max}$), we can use Eq. (17) to determine N_0 corresponding to any observed N. Accurate results are not easy to obtain because each measurement of N made in a search for the value of $N_{\max}$ has a probable relative error of $0.67/N^{\frac{1}{2}}_{\max}$. To get a value with a probable relative error of less than 0.67 percent, one must observe over 10,000 counts.

I. Counting losses: scale of s

A scale-of-s circuit feeds into the impulse counter every sth pulse it receives. If we write $s = n+1$, we see that it skips n pulses in a row and feeds in the $(n+1)$th. A perfect impulse counter would record N counts, where N is related to N_0, the number of pulses fed into the scaling circuit, by the relation $N = N_0/s$. However, if the length of an interval between the

pulses delivered by the scaling circuit (which can be called an s-tuple interval to distinguish it when necessary from an interval between the original pulses) should happen to be less than $t = k$, then a count would be missed. Turning back to Fig. 19 we recall that $y_n(t)$ is the chance that any number of pulses between 0 and n will arrive within an interval t. The chance that $n+1(=s)$ or more will arrive within t is $1-y_n(t)$, and so the chance that the sth pulse will not arrive within t is $y_n(t)$. If this interval t is equal to, or greater than, k, then the s-tuple interval will be recorded. Hence the number of s-tuple intervals recorded will be

$$N = (N_0/s)y_n(k)$$

and the "inferred number of counts," Ns, is given by

$$Ns = N_0 y_n(k). \qquad (18)$$

Figures 20 and 21 have been drawn to show how increasing the scaling factor s makes Ns, the "inferred number of counts," depart less and less from N_0, the true number of counts. The graphs are constructed as follows. For $t = k$, and remembering that $N_0 = \lambda$, Eq. (13) becomes

$$y_n(k) = \exp(-N_0 k) \\ \times [1 + (N_0 k)/1! + \cdots (N_0 k)^n/n!], \quad (19)$$

where $y_n(k)$ can be evaluated with the help of tables[19] in which the independent variable is to be used for our product $N_0 k$. The next step is to assign an arbitrary value to k, such as 0.01 sec, which is of the order of magnitude of the response time of the Cenco impulse counter, and then compute $N_0 y_n(k)$ in Eq. (18) for various values of N_0. This gives us Ns, the inferred number of counts, as a function of N_0, the true number of pulses arriving. The results are then plotted as in Figs. 20 and 21. (It is not difficult to see that the same curves could be relabeled in terms of Ns and N_0 to take care of other values of k such as 0.1 sec and 0.001 sec.) It is evident from these curves that, by increasing the scaling factor, the inferred number of counts Ns becomes more and more nearly equal to N_0, the number of pulses arriving.

With a sufficiently large scaling factor the number of counts missed by the impulse counter can be made negligibly small. The counting system still cannot count at an unlimited rate

but the losses no longer originate in the impulse counter. As scaling circuits generally have very fast times of response, the losses occur either in the G-M tube itself or in the associated control circuit.[23]

J. Effect of scaling on the distribution of intervals

The preceding section completes the theory of the effect of scaling on the counting losses in a recording device such as the Cenco impulse counter. It is instructive, however, to consider the action of a scaling circuit from a somewhat different angle. We wish now to show that the intervals (that is, the s-tuple intervals) fed into the impulse counter become more and more nearly equal as s is increased.

As in Section I, the chance that the sth pulse will not arrive within a time t is $y_n(t)$. The chance that it will not arrive within $t+dt$ is $y_n(t+dt)$. Therefore, the chance that it will arrive between t and $t+dt$ is given by

$$p_n(t)dt = y_n(t) - y_n(t+dt),$$

from which it follows that

$$p_n(t) = -(d/dt)y_n(t). \qquad (20)$$

Substituting for $y_n(t)$ from Eq. (12), we find, after some reduction, that

$$p_n(t) = \lambda(\lambda t)^n \exp(-\lambda t)/n!, \qquad (21)$$

for the chance of the sth pulse occurring between t and $t+dt$. Since $p_n(t)$ is identical with $W_n(t)$, except for the constant factor λ, Eq. (7) gives the location of its maximum, which, because of the fact that $p_n(t)$ is the gradient of $y_n(t)$, occurs at the point where $y_n(t)$ is steepest. In order to show how $p_n(t)$ [or, for that matter, $W_n(t)$] tends to concentrate closer to the maximum as n is increased, we have plotted in Fig. 17 the values of $p_3(t)$ and $p_{15}(t)$, first, to the same scale and, second, with $p_{15}(t)$ adjusted so that its maximum coincides with that of $p_3(t)$; $p_3(t)$ and $p_{15}(t)$ were chosen because they correspond to the frequently used scaling factors of 4 and 16, respectively. Fig. 17 thus gives us an alternative picture of why, with large scaling factors, few counts are lost. This is because the number of s-tuple intervals of length less than $t = k$ becomes

[23] H. Lifschutz and O. S. Duffendack, Phys. Rev. **54**, 714 (1938).

rapidly smaller with increasing n. To illustrate this, we have erected a vertical line at $\lambda t = 1$. The intercepts of the different curves on this line tell us the fraction of the intervals that are too short to register. Thus for the scales of 1, 4 and 16, corresponding to $p_0(t)$, $p_3(t)$ and $p_{15}(t)$ [or to $W_0(t)$, $W_3(t)$ and $W_{15}(t)$], respectively, we see that the percentage of intervals too short to register are 37, 6 and less than 0.1, respectively.

K. Concluding remarks

Only one reaction time, that of the Cenco counter and its associated recording circuit, is considered quantitatively in the foregoing discussion. In a more detailed analysis, it would be necessary to recognize two reaction times, the first being associated with the recording circuit and mechanical counter, and denoted (as in the preceding discussion) by k, and the second associated with the G-M tube and its control circuit and denoted by σ. Times k and σ are Types I and II reaction times, respectively. The existence of the reaction time σ modifies Eq. (18) for the relation between the inferred number of counts Ns and the number of pulses delivered to the recording circuit, N_0, into

$$Ns = \{N_0/(1 + N_0\sigma)\} y_n(k - s\sigma),$$

as long as k exceeds $s\sigma$. When $s\sigma$ exceeds k, the formula becomes

$$Ns = N_0/(1 + N_0\sigma).$$

Lifschutz and Duffendack give an illustration of a typical situation in which k is 0.015 sec and σ is 0.0005 sec. A scale of s equal to 32 would make the latter the governing factor, which means that the reaction time of the Cenco counter has no effect whatever on the losses. The losses now occur entirely ahead of the scaling circuit.

Selected References

Although several hundred papers on G-M counters, their associated circuits and theoretical questions raised by them could readily be cited, the following list is purposely limited to a few articles that should be most helpful to those who wish further information on some of the topics discussed in the present paper.

J. Strong, *Procedures in Experimental Physics* (Prentice-Hall, 1938). Contains an excellent article by H. V. Neher on counters and their associated circuits.

Duffendack, Lifschutz and Slawsky, Phys. Rev. **52**, 1231 (1937). Describes the Michigan technic for making counters.

Lifschutz and Duffendack, Phys. Rev. **54**, 714 (1938). A thoroughgoing and illuminating study, chiefly experimental, of the nature of counting losses in various parts of a typical counting unit.

Reports on Progress in Physics (London, 1937). Wynn-Williams, to whom the first scaling circuit is due, gives a summary of work in this field.

Ruark and Brammer, Phys. Rev. **52**, 322 (1937); Alaoglu and Smith, Phys. Rev. **53**, 832 (1938). Theoretical discussions of counters, losses and scaling.

Electrolytic Cells

MERLE RANDALL

Department of Chemistry, University of California, Berkeley, California[*]

THE earliest recorded observation of electrolytic action apparently is that of Sulzer.[1] He recognized a taste similar to that of ferrous sulfate when he touched his tongue to strips of lead and silver which were in contact.[2] When he touched the strips separately to his tongue no taste was apparent. The significance of this observation was lost, and it was not until the classic experiments of Galvani[3] the physician of Bologna, and of Volta,[4] professor in the University of Pavia, that electrolytic phenomena began to play a definite role in scientific thought.

Galvani placed brass in contact with a nerve and iron in contact with muscle fiber of a frog, and observed a convulsive kick when the two metals were brought together. He imagined the effect to be due to electricity generated in the frog's leg itself. Volta took the view that the electricity arose from the two different metals and not from the nerve and muscle, and built

[1] Sulzer, Historie (Mém), Acad. Roy. Sci. Belles Lettres, Berlin, **8**, 350 (1752).

[2] "Si l'on joint deux pièces, une de plomb, & l'autre d'argent, de forte que les deux bords fassent un même plan, & qu'on les approche sur la langue on en sentira quelque goût, assez approchant au goût du Vitriol de fer, pendant que chaque pièce à part ne donne aucune trace de ce goût."

[3] Galvani, Bononiensi Sci. Art. Atque Acad. Comentarii **7**, 363 (1791).

[4] Volta, Trans. Roy. Soc. London **83**, 10 (1793); **90**, 403 (1800); Phil. Mag. (4) **59**, 163, 306 (1799).

the first electrolytic cell[4] to convert chemical energy into electric energy.

PRIMARY AND SECONDARY CELLS

A cell is a mechanism in which a chemical process of the nature of a simultaneous oxidation and reduction takes place. In the cell, chemical substances disappear and others appear with either the production or utilization of electric energy.

A *primary cell* is arranged to provide a more or less efficient conversion of the free energy of a chemical process into electric energy without regard to the possibility of reversing the process, that is, of reobtaining the original chemical substances, or of restoring the cell with its original details of construction. A primary cell is used only once. An example is the ordinary dry cell, which is thrown away when the chemical process within it has run down.

A *secondary cell*, on the other hand, is so designed that the chemical process may be reversed by the expenditure of electric energy. The original substances are not only reobtained but they are restored in substantially the same physical arrangement that existed when the cell operated to convert chemical energy into electric energy. Examples are the ordinary lead-acid storage cell, the Edison alkaline storage cell, etc.

In another group of electrolytic cells, the function is the utilization of electric energy for obtaining chemical substances. There are two divisions of this group. In the first a substance, usually a metal, is dissolved at one electrode and is redeposited at the other, the electrolyte remaining substantially constant; examples are the copper refining cell, in which impure copper anodes are dissolved electrolytically in a bath of acid copper sulfate and pure copper is deposited at the cathode, and the various plating baths in which metal anodes are dissolved and corresponding metallic protective or decorative coatings are deposited on articles that serve as the cathode. In the second division of this group of cells, one (or both) of the electrodes is inert and the process functions to convert a portion of the electrolyte into other chemical substances. For example: (1) an acid copper sulfate solution is electrolyzed with insoluble anodes, copper is deposited at the cathode, sulfuric acid is regenerated and oxygen is evolved at the anode; (2) an aqueous solution of salt is electrolyzed in a diaphragm cell, sodium hydroxide and hydrogen being produced at the cathode and chlorine at the anode; (3) a solution of sodium hydroxide is electrolyzed between nickel-plated iron electrodes, hydrogen being produced at the cathode and oxygen at the anode.

Another group of cells, those generally known as *normal elements*, are so constructed that while functioning as primary cells, they are capable of giving a reproducible, reversible electromotive force. Their chief functions are to serve as standards for the International volt and as standards for measurement.

The foregoing groups of cells are of great technical importance. Many electrolytic cells, however, are used merely for measurement because the electromotive force is one of the most accurate measures of the tendency of the process to take place.

This paper will be largely concerned with a discussion of secondary cells since the principles here developed for them can be applied with obvious modifications to any of the other types of cells.

THE ELECTRON AS A CHEMICAL ELEMENT

We may greatly simplify our treatment of electrolytic cells if we consider the electron as one of the chemical elements. The elementary particle of electricity, the electron, may be regarded as an element of nuclear mass 0, atomic
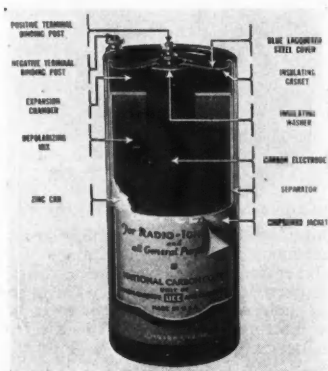

FIG. 1. Cutaway view of a dry cell, a typical primary cell. The zinc can corrodes as a part of the chemical process. Since the zinc ion which is formed diffuses into the oxidizing mix, the can is not restored on reversing the direction of the current in the cell. [Courtesy National Carbon Co., Inc.]

FIG. 2. The Daniell or "telegraph" cell. It consists of an upper zinc electrode in a saturated zinc sulfate solution and a copper electrode in the lower copper sulfate solution. The density of the latter is greater than that of the zinc sulfate solution. In this "gravity" cell the slow rate of diffusion is the mechanical barrier to internal short-circuiting. Some sort of mechanical or other barrier is essential to a cell of useful life, otherwise the cell contents come to equilibrium with the electromotive force equal to zero.
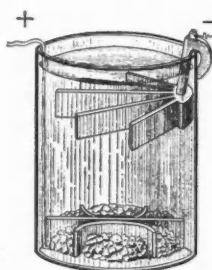
number 0 and electric charge −1. A *mol of electricity* is that amount of electricity which combines with 1 gram atom of chlorine to form 1 mol of chloride ion.

If a solution of silver nitrate is placed in a vessel containing silver electrodes, 0.00111800 gm of silver[5] will be deposited on the cathode and the same amount of silver will be dissolved from the anode for each international ampere second (coulomb) of electricity that passes. A mol of electricity,[5] 96,489 absolute coulombs, will deposit 1 mol of silver. This number is known as the *Faraday equivalent* or *faraday*. The molecular weight of electron[5] is 5.454×10^{-4}.

From this point of view chloride ion is a compound of atomic chlorine and electron. The mass of a gram atom of chlorine (6.064×10^{23} atoms) is less than the mass of a mol of chloride ion by 5.454×10^{-4} gm, an amount equivalent to the mass of a gram atom of electron.[5] We will therefore speak of electron as an element in the same sense that we speak of silver. A mol of silver ion corresponds to a gram atom of silver less a gram atom of electron. We think here of the symbol $2E^-$ as representing 2 gram atoms or 2 mols of electron rather than the charge on two individual electrons.

REACTIONS INVOLVING ELECTRON: OXIDATION AND REDUCTION

The chemist looks upon the process of simultaneous oxidation and reduction as an exchange of valence electrons by the substances oxidized and reduced.[6] By convention he writes a process

in such a way that the substances which disappear are written in the left-hand member of the equation representing the process and the substances which appear are written in the right-hand member. When a substance is *oxidized* we find that a number of mols of electron corresponding to the increase of valence stages in the process appear in the right-hand member of the equation. Thus, when ferrous ion is converted into ferric ion the essential process is

$$Fe^{++} = Fe^{+++} + E^-. \qquad (1)$$
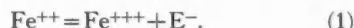
If the system is to remain electrically neutral,[7] a process of reduction must simultaneously take place. If the oxidizing agent is chlorine gas, the process will be written

$$\tfrac{1}{2}Cl_2(g) + E^- = Cl^-. \qquad (2)$$
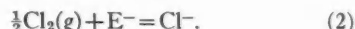
Electron therefore disappears when a substance is *reduced*. The sum of these reactions,

$$Fe^{++} + \tfrac{1}{2}Cl_2(g) = Fe^{+++} + Cl^-, \qquad (3)$$
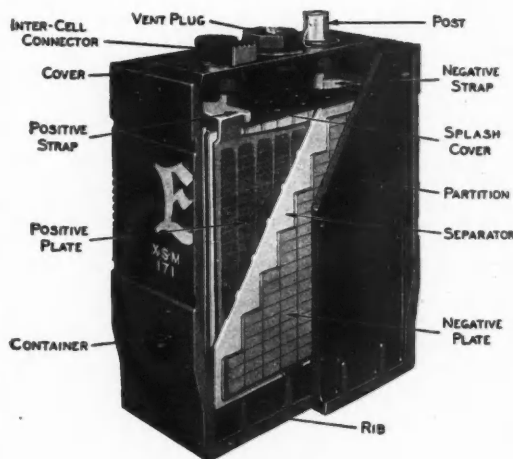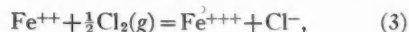
FIG. 3. Cutaway view of an automobile storage battery. The lead antimony alloy grids, or frames of the plates, keep the solid reactants in place. The separators prevent any loosened particles from straying to the opposite plate. Particles find their way to the sediment space between the ribs at the bottom. [Courtesy The Electric Storage Battery Co.]

Lewis and Randall, *Thermodynamics and the Free Energy of Chemical Substances* (McGraw Hill, 1923); (b) Randall and Young, *Elementary Applications of Physical Chemistry* (In preparation).

[7] A sphere containing 4.189 l of sodium chloride with an excess of 0.0001 mol of chloride ion would behave as a spherical condenser with a potential of 87×10^{10} v. (Ref. 7b.)

[5] Birge, Phys. Rev. Suppl. **1**, 1 (1929).

[6] For a fuller discussion of the definitions, conventions, nomenclature and derivations used in this paper see (a)

FIG. 4. An unsaturated standard cell with low temperature coefficient. The unsaturated type of cadmium cell is especially adapted to student use because of its very low temperature coefficient. In order to maintain the constancy of any standard cell, currents exceeding 0.0001 amp should never be taken from it. [Courtesy Eppley Laboratory, Inc.]

represents the total change taking place in the system during the oxidation and reduction.

THE REVERSIBILITY OF A CELL

In a previous article[8] we discussed electromotive force as a criterion for the spontaneity of a given chemical process. Many primary cells, notably the normal elements, are reversible to the extent that one is able to obtain a reproducible electromotive force when the current taken from the cell is very low. The distinction between primary and secondary cells is not a distinction based upon any difference in theory of their action, but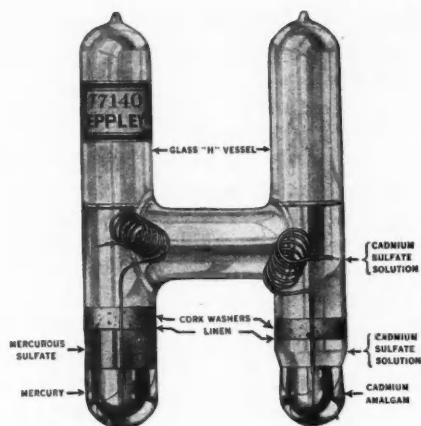 is primarily a mechanical one based upon the degree to which the original structure of the cell may be reproduced by reversing the current in the cell. For example, an ordinary run-down dry cell may be reversed, in some measure at least, by the application of an outside potential larger than the electromotive force of the process responsible for the chemical action of the cell. But this restoration of the dry cell is not very effective because there are too many side reactions and because the materials are not restored to their original locations in the cell. To understand the theory of primary cells, we must discuss thoroughly the action of a secondary cell.

[8] Randall, Am. Phys. Teacher **6**, 291 (1938).

THE REACTIONS IN THE LEAD-ACID STORAGE CELL

Let us consider the chemical reactions that take place at the electrodes of a lead-acid storage cell on discharge. At the so-called negative electrode or plate, 1 mol of lead and 1 mol of sulfate ion form 1 mol of lead sulfate and 2 mols of electricity, or

$$Pb(s) + SO_4^{--}(xM) = PbSO_4(s) + 2E^-. \quad (4)$$
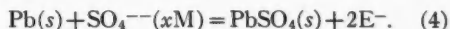
At the positive electrode, the reaction is

$$PbO_2(s) + 4H^+(2xM) + SO_4^{--}(xM) + 2E^-$$
$$= PbSO_4(s) + 2H_2O(\text{in } xMH_2SO_4). \quad (5)$$
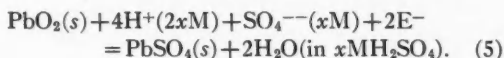
Eqs. (4) and (5) are called the *single electrode reactions* or, more frequently, the *half-reactions*. If Eqs. (4) and (5) are added, the resulting equation,

$$Pb(s) + PbO_2(s) + 2H_2SO_4(xM)$$
$$= 2PbSO_4(s) + 2H_2O(\text{in } xMH_2SO_4), \quad (6)$$
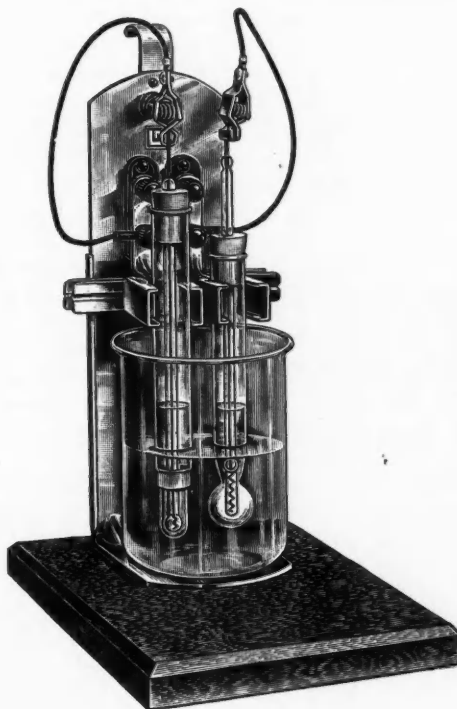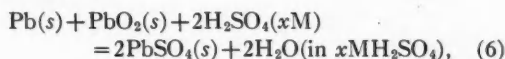
FIG. 5. A glass electrode (right) and a calomel electrode (left) mounted together in a convenient adjustable holder as used for the laboratory measurement of the pH of the electrolyte in the beaker. [Courtesy Leeds & Northrup Co.]

represents the chemical reaction for the whole cell. We should, therefore, expect the chemical reaction indicated in Eq. (6) to be the process which has taken place when 2 mols of electricity have passed through the cell. We note that it is impossible to determine, *a priori*, the *number of mols of electricity* passing through the cell, corresponding to the amount of reactants in Eq. (6), until the half-reactions have been written. If these have been written correctly, their sum is equivalent to the equation for the total reaction in the cell, and the mols of electricity expressed in the half-reactions cancel. The half-reactions also show what substances must be present in the immediate vicinity of the electrodes in order to have equilibrium.

These reactions of the lead-acid storage cell are diagrammatically illustrated[6b] in Fig. 6, which represents four stages of the process. Diagram (a) shows the cell when fully charged; (b) represents the intermediate stage of the cell while discharging in accordance with the reaction of Eq. (6); (c) represents the condition of the cell when fully discharged; and (d) shows the condition of the cell while charging in accordance with the reversed Eq. (6). At the end of the charging period the cycle is complete and the cell is again in the condition shown in (a).

The rating of any cell, primary or secondary, is expressed in ampere hours and is determined by the weight of active substance that can be converted into electricity before the potential of the cell has fallen below a useful limit. In the case of the lead-acid storage cell, its capacity is determined by the number of mols of lead and lead peroxide that can be rapidly converted, per cycle, into lead sulfate and water.

DETERMINATION OF THE PROCESS IN A CELL

The electromotive force of a cell[8] measures the tendency of some process to take place. If the process which is responsible for the measured electromotive force is reversible, the value of the electromotive force will depend only upon the activities of the substances entering into the process. Thus, the particular process responsible for the electromotive force may be identified by systematically varying the activities of the several substances which might conceivably enter into the process. Ordinarily these changes are restricted to the pressure, the temperature, and the partial pressure and concentration of the individual substances. We will consider first those changes concerned with the composition when the temperature, pressure and all variables other than partial pressure or composition of the cell remain fixed. The electrolyte in most cells is water, so we will confine our attention at the outset to cells with aqueous electrolytes.

In the article on electromotive force[8] we discussed the generalized process at constant temperature,

$$lL + mM + nE^- + \cdots = rR + sS + \cdots. \quad (8)$$

We also arbitrarily defined an *activity quotient* Q by the expression

$$Q = a_R{}^r a_S{}^s \cdots / a_L{}^l a_M{}^m \cdots. \quad (9)$$

We further indicated that the reversible electromotive force of a particular cell, containing substances with the activities a given in Eq. (9), was related to a standard electromotive force in accordance with the equation

$$\mathbf{E} = \mathbf{E}^0 - (\mathbf{R}T/n\mathbf{F}) \ln Q. \quad (10)$$

The standard electromotive force and the electromotive force are seen to be equal when the activity quotient Q is unity. The standard electromotive force more specifically measures the tendency of the process when all the reactants are at unit activity. To define the standard electromotive force we must select some con-
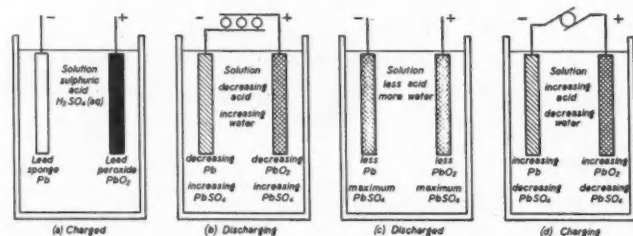

FIG. 6. The cycle of processes in a lead-acid storage cell. The state of charge is ordinarily determined by taking a hydrometer reading of the electrolyte. In automobile batteries the following specific gravity readings indicate the different charge values: 1.280, full charged; 1.225, half charged; 1.150, discharged.

venient state of each substance and arbitrarily assign unit activity to that state.

The electromotive force is such an accurate measure of the tendency of the process to take place that it furnishes one of the best available means of determining the deviation of the properties of dissolved substances from the law of perfect solutions. In order to have a basis for tabulating the individual characteristics of the several dissolved substances, it is necessary to choose a standard state or standard of reference for each substance. For pure solids or pure liquids at atmospheric pressure, the activity is assumed to be unity at all temperatures. Its variation with pressure will be discussed later.

The Standard State of a Solute

In the case of solutions, it is convenient to define the standard state of the solute and solvent with reference to the same solution. Because of the extreme simplicity of the relations existing in an infinitely dilute solution we make this solution the basis of our definition. By definition, the activity of any substance in its standard state is unity. From a practical standpoint we wish to express the activity and the molality of an ion on the same basis and in such a way that for ordinary purposes the activity may be taken equal to the molality.[9] We thus choose the standard state of an ion so that $a_+/m_+ = 1$ when $m_+ = 0$. These considerations are equivalent to the definition of the standard state as a hypothetical one having unit activity and all the other thermodynamic properties of an infinitely dilute solution.

Actual solutions in which an ion may possess unit activity are not solutions of the ion in its standard state. The activity of such an ion initially at unit activity would not remain at unit activity but would change with temperature and with pressure because the partial molal heat content, partial molal volume, etc., of the solute ion are not those of the ion in an infinitely dilute solution. Such definitions are convenient for strong electrolytes, which may be regarded as completely dissociated.

We find in the case of electrolytes that, while

[9] In the infinitely dilute solution Henry's law may be stated in the form: The fugacity is proportional to the molality in an infinitely dilute solution.

the activity of an ion is equal to its molality m in the infinitely dilute solution, this is not true for actual solutions. We may, however, always determine the value of the activity in any solution by multiplying the molality of an ion by a number known as its activity coefficient. This activity coefficient enables us to tabulate all of the special properties of individual ions and electrolytes.

We shall not discuss the activity of individual ions. As a matter of fact, it is impossible to define uniquely an individual ion activity. In every device so far attempted, it has been found impossible to measure single ion activities. We can measure only their products or their quotients. If we speak of the activity of an ion, therefore, we must remember that we refer to this activity in the sense that the positive and negative activities are to be multiplied as in Eq. (11). We always obtain a mean activity coefficient of the ions of some salt. For example, if an electrolyte dissociates into $\nu(=\nu_+ + \nu_-)$ ions according to the equation

$$X = \nu_+ X^+ + \nu_- X^-,$$

we may write for the mean activity of the ions of the salt X,

$$a_+{}^\nu + a_-{}^\nu = a_2; \quad a_\pm = (a_2)^{1/\nu}. \tag{11}$$

Likewise, we may define a mean molality of the salt X by the equation

$$m_\pm = m(\nu_+{}^\nu + \nu_-{}^\nu)^{1/\nu}. \tag{12}$$

We may now write

$$\gamma = a_\pm/m_\pm = (\gamma_+{}^\nu + \gamma_-{}^\nu)^{1/\nu}, \tag{13}$$

where γ is the mean activity coefficient of the ions of the salt. For the sake of consistency of nomenclature we might have written the activity coefficient as $\gamma_\pm$. Having established these conventions, as well as the similar convention that the activity of an undissociated substance approaches its molality at infinite dilution, we are now in a position to study the way in which the activity coefficients of various substances change with the concentration.

The Activity Coefficient

It is beyond the scope of our discussion to go into the modern theory of electrolytes. In order

to determine the character of the chemical process, we must, however, examine the way in which the activity coefficients of typical electrolytes vary with concentration, not only in their pure solutions but also in mixtures with various electrolytes.

Raoult's law, in its most general form,

$$f_1 = f_1^0 N_1, \qquad (14)$$

states that the escaping tendency of a substance is that of the pure solvent times its mol fraction. It is a limiting law which describes the behavior of the solvent in very dilute solutions. It also quite accurately represents the behavior of some mixtures within certain concentration ranges, namely, for the solvent in very dilute solutions.

Henry's law substitutes an empirical constant k_2 in Raoult's law in place of f_2^0, the escaping tendency of the pure solute. Expressed in the form,

$$f_2 = k_2 N_2, \qquad (15)$$

Henry's law is found to give a satisfactory representation of the activity-concentration relations of the solute in very dilute solutions of non-electrolytes.

Modern ionic theory postulates that a strong electrolyte is completely dissociated. Of course, the pure thermodynamist need not be interested in this phenomenon, but the fact remains that Eq. (15) is approximately valid for each of the ions of a dissociated electrolyte. Therefore, our equations relating to the activity of strong electrolytes become far simpler when we postulate this mechanism of complete dissociation.

In dealing with the activity coefficient, it will perhaps be easier to utilize an equation analogous to Eq. (10), namely,

$$\bar{F}_i = \bar{F}_i^0 + RT \ln a_i, \qquad (16)$$

which is in reality the equation by which the activity is defined. In Eq. (16), $\bar{F}_i$ is the partial molal free energy of an ion; $\bar{F}_i^0$ is its molal free energy in the standard state or at unit activity. The molal free energy, like the electromotive force, is an intensive property in that it is the free energy per mol and therefore independent of the amount of substance involved. The free energy, however, is an extensive quantity analogous in this respect to heat content, energy
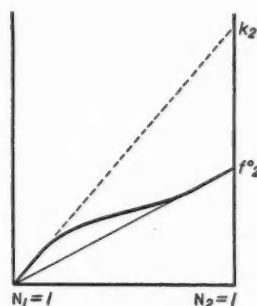
FIG. 7. The constant of Henry's law. The fugacity of a solute always approaches some straight line in the case of the dilute solution. The extrapolated intercept of this line upon the axis of pure solute is the constant of Henry's law. It represents the fugacity of a fictitious pure solute that has the same properties as in the very dilute solution. As we come to solutions more nearly perfect over the entire range, the properties of this fictitious pure solute become more like those of the actual pure solute.

content, etc. Although it is an extensive property, it is not practical to assign an absolute value to the free energy. This inability to assign an absolute value, however, does not limit the usefulness of the function, for the free energy content may be thought of as the net work content. Since we must always start with substances in one state and finish with them in another, the net work obtainable from a process, for example, that from the generalized process in Eq. (8), is

$$\Delta F = F_{II} - F_I = r\bar{F}_R + s F_S + \cdots$$
$$- (l\bar{F}_L + m\bar{F}_M + \cdots). \qquad (17)$$

The electromotive force is also a measure of the net work, per mol of electron transferred, available from a given process, say, that of Eq. (8). Hence, if n is the number of equivalents or number of mols of electron concerned in the process, then $\Delta F = -n\mathbf{E}\mathbf{F}$, where $\mathbf{F}$, the Faraday equivalent, is measured in calories per equivalent or[5] 23,054 cal/mol of electron.

Now in dilute solutions the molality (mols per thousand grams of water) is nearly proportional to the mol fraction $m(=55.51 N_2)$ and, in accordance with Eq. (15), we might substitute the molality, m_i, for the activity of the ion, a_i, in Eq. (16). However, the standard state in which the molal free energy is $\bar{F}_i^0$ was defined in such a way that the positively and negatively charged ions constituting the solution are situated at an

infinite distance from one another. In any solution of finite concentration, we must have an equal number of positive charges and negative charges upon the ions, otherwise the solution will be charged with respect to the surroundings.[7]

By a careful consideration of the orientation of the positive and negative charges in a finite solution (Fig. 8) Debye and Hückel[10] calculated the work necessary to bring a mol of ion from a condition of infinite attenuation up to the electrical environment of the finite solution. Their formulation is equivalent to assuming that the partial molal free energy of an ion in any moderately dilute solution is equivalent to the partial molal free energy of an uncharged ion, $\bar{F}_i{}^0 + RT \ln m_i$, the molal free energy in the standard state plus the work necessary to bring an uncharged ion from the standard state to the concentration in question, plus the work necessary to carry the charges from a condition of infinite attenuation (postulated by the standard state) up to the concentration in question. Debye and Hückel demonstrated that it was possible to calculate this electrical work in very dilute solutions and arrived at equations dependent on the dielectric constant and the absolute temperature of the solvent but devoid of empirical constants.

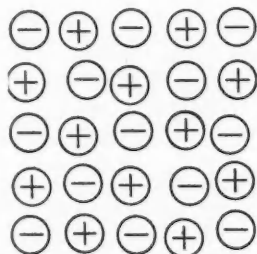Lewis and Randall[11] had already demonstrated, and the result was confirmed theoreti-

FIG. 8. Orientation of ions in a solution. In an electrolytic solution of finite strength, the tendency of each positively charged ion to attract negatively charged ions and to repel positively charged ions results in a more or less definite orientation of the ions within the electrolyte. As the solution becomes more dilute, the effect is reduced. The tendency to assume a definite lattice is opposed by the thermal agitation of the solvent molecules, and this results in the continual shifting and distortion of the lattice.

[10] Debye and Hückel, Physik. Zeits. **24**, 185 (1923); Debye, *ibid.* **25**, 97 (1924).
[11] Lewis and Randall, J. Am. Chem. Soc. **43**, 1112 (1921).

cally by Debye and Hückel,[10] that the activity coefficient of an ion was largely independent of the chemical nature of the ion, being dependent almost entirely upon the distance between the ions and the number of charges upon them. Lewis and Randall defined a quantity which they called the *ionic strength*,

$$\mu = \tfrac{1}{2}\sum m_i z_i{}^2, \tag{18}$$

where m_i is the molality of the ion and z_i is its valence or number of charges independent of their sign. The total number of positive charges must, of course, equal the total number of negative charges if the electric neutrality of the solution is maintained. It is thus seen that a bivalent ion, such as barium ion, Ba^{++}, produces four times the effect of a univalent ion, such as chloride ion, Cl^-; and that lanthanum ion, La^{+++}, is nine times as effective in producing an electrical environment as is sodium ion, Na^+.

Without going into the details of the calculation, we find that the mean activity coefficient of a salt in dilute solution is given by the equation[12]

$$\log \gamma = A z_+ z_- d^{\frac{1}{2}} \mu^{\frac{1}{2}}, \tag{19}$$

where d is the density of water and

$$A = -(2\pi)^{\frac{1}{2}} N_0{}^2 \epsilon^3 / 2.3026 R^{\frac{3}{2}} T^{\frac{3}{2}} D^{\frac{3}{2}} V^{\frac{1}{2}} \tag{20}$$

is the work term of Debye and Hückel[10] in which N_0 is the Avogadro number[5] (6.064×10^{23} mol^{-1}), ϵ is the elementary charge on the electron (4.770×10^{-10} esu), R is the gas constant (8.3136×10^7 erg deg^{-1} mol^{-1}), T is the absolute temperature, D is the dielectric constant and V is 1000 cm^3. The values of $Ad^{\frac{1}{2}}$ at several temperatures[12] are given in Table I.

If the compositions are expressed as concentrations (mols/l), then we may adopt the suggestion of Randall and Vietti[13] and define an activity coefficient in such a way that

$$a_i = m_i \gamma_i = c_i \gamma_{ci}. \tag{21}$$

The definition avoids a difficulty of Debye and Hückel, who implied as standard state a solution whose composition must vary with temperature as a result of the expansion of the solution. Randall and Vietti's suggestion is equivalent to

[12] Randall and West, to be published.
[13] Randall and Vietti, J. Am. Chem. Soc. **50**, 1526 (1928).

TABLE I. *Values of the constant A, $Ad^{\frac{1}{2}}$ and log d for calculating the limiting Slope of the curve: logarithm of the activity coefficient vs. square root of ionic strength.*

$t(°C)$	$A(DH)$ [LOG f, EQ. (20)]	$Ad^{\frac{1}{2}}$ [LOG γ, EQ. (19)]	LOG d [LOG γ, EQ. (22)]
0	−0.4882	−0.4882	−0.0001
10	.4956	.4955	.0001
20	.5042	.5038	.0008
25	.5090	.5083	.0013
30	.5139	.5128	.0019
40	.5247	.5227	.0034
50	.5365	.5333	.0052
60	.5495	.5449	.0073
70	.5634	.5572	.0097
80	.5785	.5703	.0124
90	.5948	.5844	.0153
100	.6112	.5984	.0185

the expression

$$\log \gamma_c = A z_+ z_- \mu_c^{\frac{1}{2}} - \log d, \qquad (22)$$

where

$$\mu_c = \frac{1}{2}\sum c_i z_i^2. \qquad (23)$$

The values of log d for various temperatures are given in the fourth column of Table I.

Randall and Vietti[13] plotted the quotient $\log \gamma / z_+ z_-$ as a function of the square root of the ionic strength (Fig. 9). They found that the curves for various typical strong electrolytes appeared as a family of curves asymptotic to a straight line whose slope was $Ad^{\frac{1}{2}}$. In examining Fig. 9, we find that the activity coefficients of some substances, such as the halogen acids, are much higher even in moderately dilute solutions than is predicted by the simple postulates of Debye and Hückel. We also find that the substances arrange themselves in a very definite order characteristic of the pairs of ions making up the salt. We also note that the curves for a few substances, for example, cadmium chloride and sulfuric acid, fall below the 45° slope in the dilute range but cross over and are, in general, higher than the

45° slope in solutions of moderate concentrations. This result may be expressed in either of two ways. For example, Gronwall, La Mer and Sandved[14] have shown that, by taking account of higher terms and introducing empirical parameters, they could predict the trend of the curve log γ vs. $\mu^{\frac{1}{2}}$ with considerable certainty. Bjerrum[15] postulates that there is an association of the ions in moderately concentrated solutions, or, in the case of weak acids, he assumes that there is a fraction only of the dissolved substance dissociated.[16] All of these assumptions must, of course, yield the same value for the difference in the free energy of the solute in its measured state and in the standard state, if the correct account of the various factors is taken. The observations summarized in Fig. 9 are of great importance in determining the nature of the processes taking place in an electrolytic cell.

[14] Gronwall, La Mer and Sandved, Physik. Zeits. **29**, 358 (1928).
[15] Bjerrum, Det. Kgl. Danske Videnskab. Selskab. Math.-fys. Medd. **7**, No. 9 (1926).
[16] (a) Randall and Failey, Chem. Rev. **4**, 271, 285, 391 (1927); (b) Randall and Allen, J. Am. Chem. Soc. **52**, 1814 (1930).

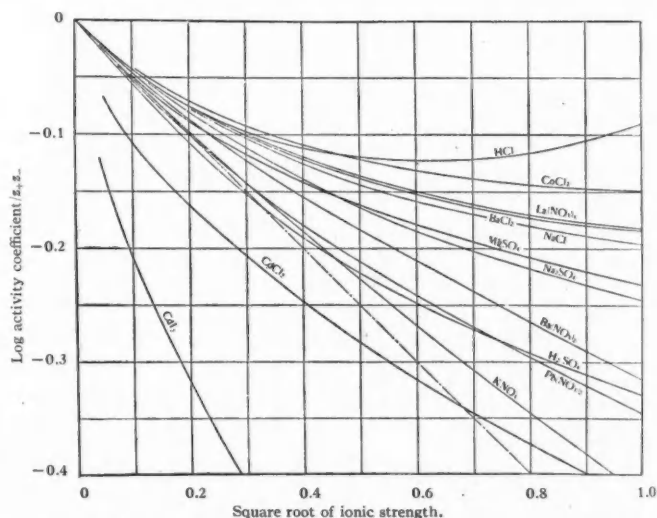
FIG. 9. Activity coefficients of aqueous solutions of electrolytes. The quotient of the mean logarithm of the activity coefficient and the product of the valencies of the ions is plotted against the square root of the ionic strength. The plot is usually made on a standard scale of 0.1 unit in log $\gamma/z_1 z_- = 10$ cm and 0.2 unit in $\mu^{\frac{1}{2}} = 10$ cm. [Courtesy J. Am. Chem. Soc.]

## THE ACTIVITY COEFFICIENT AND THE VARIATION OF THE ELECTROMOTIVE FORCE

Consider for purposes of illustration the electromotive force of the hydrogen-silver chloride-silver cell. Such a cell consists of two half-cells in which the reaction or cell process taking place in the silver-silver chloride half-cell is

$$AgCl(s) + E^- = Ag(s) + Cl^-. \quad (24)$$
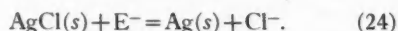
In the hydrogen half-cell the reaction is

$$\tfrac{1}{2}H_2(g) = H^+ + E^-. \quad (25)$$
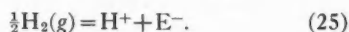
Their sum gives the reaction for the complete cell,

$$AgCl(s) + \tfrac{1}{2}H_2 = Ag(s) + H^+ + Cl^-. \quad (26)$$
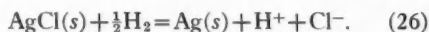
The reversible potential of this cell, as measured by Randall and Young[17] at 25°C and for several molalities, is given in the second column of Table II. The pressure of the hydrogen in each case was 1 atmos. It is noted that for concentrations of hydrochloric acid above approximately $9M$ the electromotive force is negative. In other words, hydrogen gas under 1 atmos tends spontaneously to reduce solid silver chloride to metallic silver, hydrogen ion and chloride ion below $9M$. Above this molality the electromotive force is negative, or metallic silver tends to dissolve in hydrochloric acid to form silver chloride and hydrogen.

Referring to Eq. (10) and noting that n is 1, we find

$$\mathbf{E} = \mathbf{E}^0 - (RT/nF) \ln a(H^+) a(Cl^-). \quad (27)$$

Now substituting the molality times activity coefficient $m\gamma$ for the activity of the ion and rearranging, we find

$$\log \gamma - \mathbf{E}^0/(2 \times 2.3026 \times RT/nF)$$
$$= -\mathbf{E}/(2 \times 2.3026 RT/nF) - \log m. \quad (28)$$

Substituting the values of RT and F, we have

$$\log \gamma - \mathbf{E}^0_{298.1}/0.11828$$
$$= -\mathbf{E}_{298.1}/0.11828 - \log m. \quad (29)$$

In the case of hydrochloric acid the ionic strength is equal to the molality [Eq. (18)]. Whence we may plot (Fig. 10) the value of $(\log \gamma - \mathbf{E}^0/0.11828)$ or $(-\mathbf{E}/0.11828 - \log m)$ against the square root of $m(=\mu^{\frac{1}{2}})$ as in Fig. 9. If the plot

[17] Randall and Young, J. Am. Chem. Soc. **50**, 989 (1928).

TABLE II. *Potentials of the hydrogen-silver chloride-silver cell at 25°C.*

m	$\mathbf{E}_{298.1}$	LOG γ +C	$m^{\frac{1}{2}}$	LOG γ
0.1015	0.3514	−1.9772	0.3186	−0.0995
.0972	.3536	−1.9772	.3118	−.0995
.0974	.3539	−1.9807	.3121	−.1030
.1233	.3422	−1.9841	.3512	−.1064
4.095	.1194	−1.6218	2.0235	.2559
14.84	−0.0953	−0.3657	3.8530	1.5120

is made on translucent paper, we find that it may be superposed over the plot of Fig. 9.

Now utilizing the curve $\log \gamma$ vs. $m^{\frac{1}{2}}(=\mu^{\frac{1}{2}})$ for hydrochloric acid as a guide to extrapolating the plot of $(-\mathbf{E}/0.11828 - \log m)$ vs. $\mu^{\frac{1}{2}}$, we find the intercept at $\mu^{\frac{1}{2}}=0$ to be -1.8777. This is the value of $-\mathbf{E}^0/0.11828$. Whence $\mathbf{E}^0_{298.1}$, the standard electromotive force, is 0.2221 v. If we now substitute the value of $\mathbf{E}^0/0.11828$ back in Eq. (29) we may calculate $\log \gamma$ at the several molalities given in the last column of Table II.

Since we formulated Eq. (27) on the basis of the postulated process of Eq. (26), the fact that the activity coefficient of hydrochloric acid, as determined by the extrapolation which we have performed, was equal to the value of the activity coefficient of hydrochloric acid determined by other methods, provides strong evidence that the process assumed in Eq. (26) is correct. Another form of this criterion as to the correctness of the process assumed in Eq. (26) is the fact that the plot of $(-\mathbf{E}/0.11828) - \log m$ vs. $\mu^{\frac{1}{2}}$ superposes upon the plot of $\log \gamma$ vs. $\mu^{\frac{1}{2}}$ of hydrogen chloride. Another criterion depending upon the variation of the electromotive force of the cell with the temperature and with the pressure of the hydrogen will be discussed later.

THE HYDROGEN-CALOMEL CELL IN MIXED ELECTROLYTES

Loomis, Essex and Meacham[18] showed that the product of the hydrogen ion activity and of the chloride ion activity in a mixed electrolyte, say, hydrochloric acid mixed with sodium chloride or with sodium perchlorate, is dependent only on the total molality. If the added electrolyte is a multivalent salt such as barium

[18] Loomis, Essex and Meacham, J. Am. Chem. Soc. **39**, 1133 (1917). See Ref. 7(a)
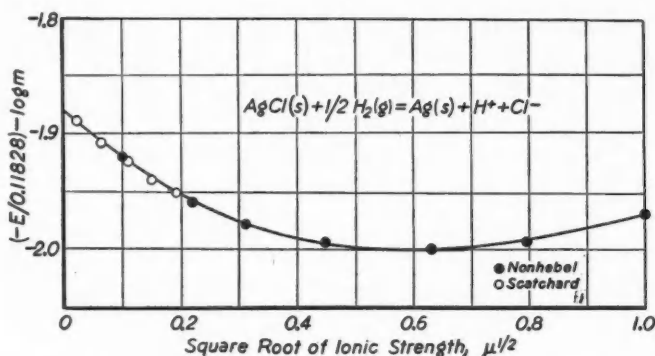
FIG. 10. Logarithm of mean activity coefficient in the hydrogen-silver chloride-silver cell. The value of the standard electromotive force divided by the conversion factor 0.11828 is obtained by superposing this curve upon Fig. 9. From Eq. (26) we see that the curve should superpose upon the activity coefficient curve for hydrogen chloride in that figure. The value of $-E°/0.11828$ is the intercept at $\mu^{\frac{1}{2}}=0$. Results of Scatchard [J. Am. Chem. Soc. **47**, 641 (1925)] and of Nonhebel [Phil. Mag. (7) **2**, 1085 (1926)] are included in the plot.

chloride, then the activity coefficient is dependent mainly upon the ionic strength as defined in Eq. (18). Randall and Breckenridge[19] measured the hydrogen-calomel cell in mixtures of hydrochloric acid with barium chloride and with lanthanum chloride. The process is

$$HgCl(s)+\tfrac{1}{2}H_2(g)=Hg^{(l)}+H^++Cl^-. \quad (30)$$

Their results are shown in Figs. 11 and 12, where $(-E/0.11828)-\log m_{\pm}$ is plotted against the square root of the ionic strength. These plots will superpose on Fig. 10.

When we come to consider a mixture of electrolytes, there is at first sight some ambiguity as to the meaning of the activity or molality of a substance such as hydrochloric acid in a mixture with other chlorides. Consider, for example, a solution made by dissolving 0.1 mol of HCl and 0.1 mol of $BaCl_2$ in 1000 gm of water. The molality of the hydrogen ion is obviously 0.1, that of the barium ion 0.1, and that of the chloride ion 0.3. The chloride ion furnished by the solution of hydrochloric acid has exactly the same properties as the chloride ion furnished by the solution of barium chloride. By Eq. (10), we find that the potential of the cell depends upon the product of the activities of the hydrogen and chloride ions and so far as

[19] Randall and Breckenridge, J. Am. Chem. Soc. **49**, 1435 (1927).

the potential of the cell is concerned, we would be no more correct in saying that the molality of the hydrochloric acid is 0.1 (the molality of the hydrogen ion) than in saying that it is 0.3 (the molality of the chloride ion). We can, however, speak of the *geometric mean molality* of the hydrochloric acid in the mixed electrolytes and define this mean molality by Eq. (12). Whence we have for the mean activity coefficient the expression that we have already given in Eq. (13). The mean molality of a pair of ions is, therefore, significant in characterizing the behavior of electrolytic cells involving a particular pair of ions.

The several curves in Fig. 11 are for the several indicated ratios of the molality of HCl and of $BaCl_2$ in the mixed electrolyte. The lowest dotted curve represents hydrochloric acid in the presence of such a large proportion of barium chloride that the electrolyte is essentially pure barium chloride. This last curve will superpose slightly above that of barium chloride in Fig. 9. This would seem to indicate that the effect of one electrolyte upon another must be of the same character as that which governs the value of the constant in Henry's law.

If Raoult's law [Eq. (14)] were valid throughout the entire range of concentration, then the constant of Henry's law [Eq. (15)] would be the same as the constant of Raoult's law, and the constant of Henry's law would be independent of the kind of solvent. In other words, the molecules of the solvent would have just the same effect upon the surrounding molecules of solute as have the solute molecules themselves. But, as might be expected when each solute molecule is surrounded by a large number of solvent molecules, the individual properties of the solute molecule play a minor role and the properties of the solvent play a major role. Hence we find the constant of Henry's law dependent upon the solvent as well as upon the solute.

Lewis and Randall[6, 11] gave an approximate rule for the ionic strength by stating that the activity coefficient of a given salt is the same in
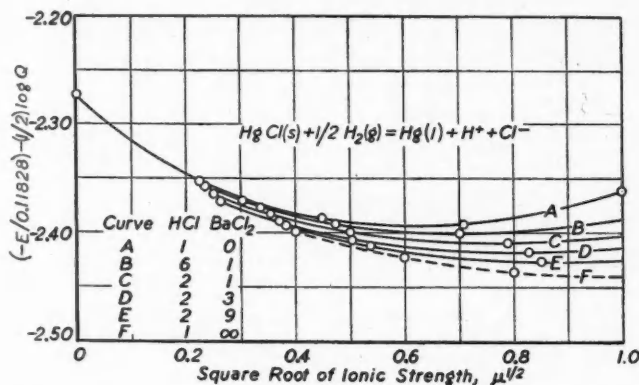
FIG. 11. Hydrogen-calomel cell with added barium chloride. The various curves represent various ratios of hydrochloric acid to barium chloride. The lowest dotted curve represents the limiting value when the ratio of hydrochloric acid to barium chloride becomes zero.

all aqueous solutions of the same ionic strength. This is analogous to the statement that the constant of Henry's law is always equal to the constant of Raoult's law. Obviously, as is evidenced by perusal of Figs. 11 and 12, when hydrogen ion is in an environment consisting largely of barium ions or of lanthanium ions, the mean activity coefficient of the hydrochloric acid has a tendency to become more nearly identical with that of barium chloride or lanthanium chloride, as the case may be.

SIDE REACTIONS

One explanation of the peculiar shape of the curve $\frac{1}{2} \log \gamma$ vs. $\mu^{\frac{1}{3}}$ for sulfuric acid in Fig. 9 is that sulfuric acid does not completely dissociate in accordance with the equation

$$H_2SO_4 \ (aq) = 2H_2^+ + SO_4^{--}, \quad (31)$$
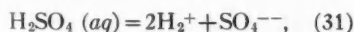
but rather that a considerable proportion of the sulfuric acid dissociates as

$$H_2SO_4 \ (aq) = H^+ + HSO_4^-. \quad (32)$$
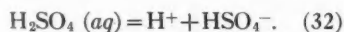
The relationship between the activity of HSO_4^- and of H^+ and SO_4^{--} is given by the equation

$$HSO_4^- = H^+ + SO_4^{--};$$
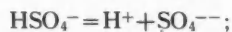
$$K_2 = a(H^+)a(SO_4^{--})/a(HSO_4^-).$$
$$(33)$$

The value of the constant K_2 in Eq. (33) can be approximately determined by a number of methods. Thus we may calculate the amount of sulfate ion, SO_4^{--}, present in a given solution.

If we assume Eq. (32) rather than Eq. (31) to be the main reaction, deduct the amount of SO_4^{--} from that of the total sulfate to obtain the HSO_4^- and deduct the amount of HSO_4^- from the total hydrogen ion constituent, we would expect the curve $\log \gamma$ vs. $\mu^{\frac{1}{3}}$ to have a shape similar to that of hydrochloric acid, that is, similar to the curve of a uni-univalent salt. Of course, sulfate ion, being bivalent, contributes four times as much to the ionic strength as does HSO_4^-, which is a univalent ion. In this way we can actually obtain a plot of the mean activity coefficient of hydrogen and hydrosulfate ions against the square root of the ionic strength that lies only slightly below the curve for hydrochloric acid.

On the other hand, if we assume that the main reaction is that of Eq. (31), and that the reaction of Eq. (32) is a side-reaction, then taking into account the proportion of hydrosulfate ion, HSO_4^-, calculated from Eq. (33), we find that the curve of $\frac{1}{2} \log \gamma$ (γ being the mean activity coefficient of the hydrogen and sulfate ions
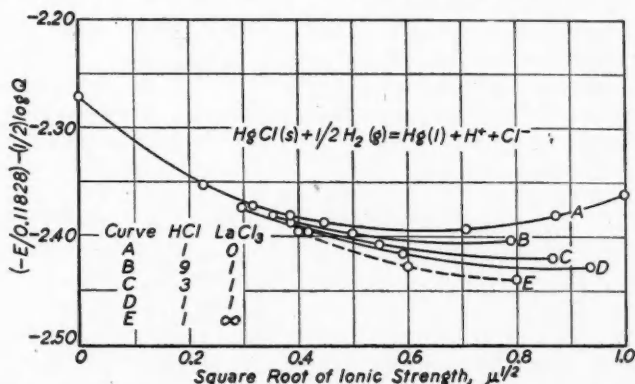
FIG. 12. Hydrogen-calomel cell with added lanthanum chloride. The effective concentration of the hydrochloric acid in the presence of added lanthanum chloride is the mean molality of the hydrogen and chloride ions. As in Fig. 11, the lowest dotted curve represents a solution with a vanishingly small proportion of hydrogen ion.

actually present) plotted against the square root of the ionic strength is nearly identical to the curve for a typical uni-bivalent salt, namely, calcium chloride.

Another illustration of the utilization of the foregoing principle is to be found in the cells studied by Bray and Hershey[20] corresponding to

$$Fe^{+++} + \tfrac{1}{2}H_2(g) = Fe^{++} + H^+. \quad (34)$$

The reaction is the reduction of ferric chloride by hydrogen gas to ferrous chloride in acid solution. They found varying degrees of hydrolysis of the ferric ion, depending upon the concentration of hydrochloric acid, and they also found varying amounts of $FeCl^{++}$ formed by another side-reaction,

$$Fe^{+++} + Cl^- = FeCl^{++}. \quad (35)$$

Randall[21] and Randall and Shaw[22] extended the idea of Randall and Vietti.[13] Empirically and in accordance with the Debye and Hückel theory they showed that when $-(1/\Delta z^2) \log K_m$ for any reaction involving ions (K_m is the equilibrium molality expression) is plotted against $\mu^{\frac{1}{2}}$ on the standard scale used by Randall and Vietti, then the curve will superpose upon one of the activity coefficient curves when the reaction is properly chosen. The value of $-(1/\Delta z^2) \log K$ is given by the intercept of the family of curves on the axis of $\mu^{\frac{1}{2}} = 0$. Such reactions may involve one or more undissociated substances, as for example, the reaction

$$CaCO_3 \text{ (solid)} + H_2CO_3 = Ca^{++} + 2HCO_3^-, \quad (36)$$
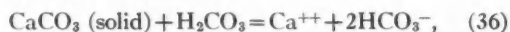
in which case $\Delta z^2 = 6$. In reaction (34) Δz^2 is -4 and in reaction (35) it is -6. If the process has been correctly formulated, the plot of $-(1/\Delta z^2) \log K_m$ vs. $\mu^{\frac{1}{2}}$ will fall properly in the family of curves of Fig. 9.

If the process is not properly formulated, and if proper allowance for side-reactions has not been made, then the plot will not follow in the family of curves as is indicated, for example, in Fig. 13. Smith,[23] in measuring the solubility of zinc carbonate in aqueous carbonic acid, did not determine the composition of the solid phase. Depending upon the composition of the solid

[20] (a) Bray and Hershey, J. Am. Chem. Soc. 56, 1889, 2799 (1934); (b) Hershey and Bray, ibid. 58, 1760 (1936).
[21] Randall, J. Chem. Ed. 8, 1062 (1931).
[22] Randall and Shaw, J. Am. Chem. Soc. 57, 427 (1935).
[23] Smith, J. Am. Chem. Soc. 40, 883 (1918).
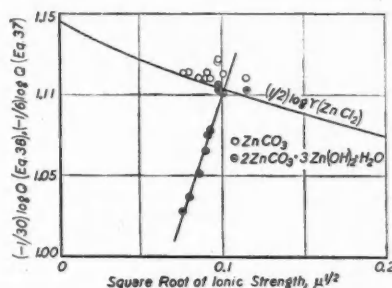
FIG. 13. Solubility of zinc carbonate in aqueous carbonic acid at 25°C. The fact that the function of log activity coefficient of zinc chloride is approximately followed when the solid phase is the normal carbonate is strong evidence of the correctness of the assumption of normal carbonate. The ions in solution correspond to aqueous zinc bicarbonate which is a salt not unlike zinc chloride. [Randall, Ref. 21.]

phase, the solubility is represented by one of the following equations:

$$ZnCO_3(s) + H_2CO_3(aq) = Zn^{++} + 2HCO_3^-, \quad (37)$$

$$2ZnCO_3 \cdot 3Zn(OH)_2 \cdot H_2O(s) + 8H_2CO_3(aq)$$
$$= 5Zn^{++} + 10HCO_3^- + 7H_2O(l). \quad (38)$$
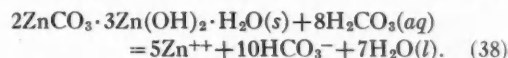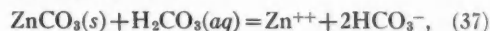
In Eq. (37), $\Delta z^2 = 6$ and in Eq. (38), $\Delta z^2 = 30$. In Fig. 13 the circles represent the values of

$$(-1/\Delta z^2) \log K_m$$
$$= -\tfrac{1}{6} \log [m(Zn^{++})m^2(HCO_3^-)/m(H_2CO_3)]$$

[Eq. (37)] and the points indicated with double circles are the values of

$$(-1/\Delta z^2) \log K_m = -1/30 \log [m^5(Zn^{++})$$
$$\times m^{10}(H_2CO_3)/m^8(H_2CO_3)]$$

[Eq. (38)]. The solid curve is the curve of $(1/z_+z_-) \log \gamma (ZnCl_2)$ plotted against the square root of the ionic strength. It is evident from the plots that the reaction formulated by Eq. (37) represents more nearly the process measured by Smith[23] than does the reaction of Eq. (38), because the points calculated on the basis of Eq. (37) fall on the zinc chloride curve. The identity of the points calculated on the basis of Eq. (37) with the curve of zinc chloride may be interpreted by saying that zinc bicarbonate and zinc chloride have almost the same activity coefficients. From other considerations we are led to believe that zinc bicarbonate will give almost the same activity coefficient curve as zinc chloride, additional proof of the validity of the choice of Eq. (37) instead of Eq. (38).

PROCESS OF THE LEAD-ACID STORAGE CELL

We have assumed that the process of the lead-acid storage cell is represented by Eq. (6). Harned and Hamer[24] made an extensive investigation of the single electrode potentials and the potentials corresponding to reactions (4) and (6) at various temperatures and with varying concentrations of sulfuric acid. Their results corresponding to Eq. (6) for 25°C are given in Table III, where the first column gives the

TABLE III. *Variation of the voltage of the lead-acid cell with molality from 0 to 7 M at 25°C.*

$m(H_2SO_4)$	E 25°	$\mu^{\frac{1}{2}}$	LOG $\gamma + C$
0.05	1.76182	0.388	4.96511
.10	1.79626	.549	5.06217
.20	1.83122	.775	5.16069
.50	1.87874	1.227	5.29461
1.00	1.91945	1.731	5.40934
1.50	1.94780	2.122	5.48923
2.00	1.97100	2.451	5.55461
3.00	2.01379	3.000	5.67520
4.00	2.05291	3.468	5.78545
5.00	2.08973	3.871	5.88921
6.00	2.12368	4.239	5.98489
7.00	2.15502	4.580	6.07321

molality of the sulfuric acid, the second the measured potential, and the last the activity coefficient.

The value of Q in Eq. (6) is

$$Q = a^2(H_2O)/(m^4(H^+)m^2(SO_4^{--})). \quad (39)$$

The value of n, the number of equivalents, is 2. The value of Δz^2 is -12. Hence, in accordance with the foregoing principles, we write

$$\log \gamma + E^0/(6 \times 0.05914)$$
$$= E/(6 \times 0.05914) + (1/12) \log Q. \quad (40)$$

Plotting the right-hand member of the equation against $\mu^{\frac{1}{2}}$, we obtain the curve in Fig. 14, which superposes upon the curve of the activity coefficient of sulfuric acid in Fig. 9 and gives a limiting value of $E^0/(6 \times 0.05914) = 5.75586$ or[24]

$$PbO_2(s) + Pb(s) + 2H_2SO_4(aq) = PbSO_4(s)$$
$$+ 2H_2O(l); E^0_{298.1} = 2.04241. \quad (41)$$
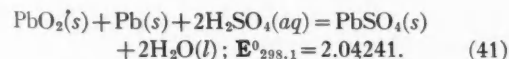
The superposition of the plot of Fig. 14 over that for sulfuric acid as in Fig. 9 gives a very strong proof of the validity of the process postulated

[24] Harned and Hamer, N. R. L. Report. No. P-1071 (Navy Dept., Bur. of Eng., Washington, 1934); J. Am. Chem. Soc. 57, 27 (1935).

in Eq. (6). This change of the voltage of the lead-acid storage cell with the concentration and with the density of the acid is shown in Table IV and Fig. 15.

EFFECT OF PRESSURE ON ELECTROMOTIVE FORCE

In the previous sections we have discussed at length the effect of the change of concentration upon the electromotive force. If we avoid irregularities in the surface conditions of metal electrodes, the activities of the various cell constituents and, hence, the electromotive force of the cell will ordinarily depend, at a given temperature, only upon pressure and concentration. The pressure effect is usually negligible except when some one of the substances concerned is a gas. In such case the activity may be found by the methods of Lewis and Randall,[6] or it frequently suffices to assume a perfect gas, and to write the activity of a gas equal to the partial pressure.

When the hydrogen electrode came into common use, it was important to show that this electrode gives a reversible electromotive force, and one which is in accord with the assumed cell reaction. The simplest test was to study the effect of pressure upon the electromotive force of a cell with hydrogen and calomel electrodes in hydrochloric acid, the cell reaction being that of Eq. (30).

If pure hydrogen passes through such a cell against a given external pressure, the partial pressure of the hydrogen is equal to the difference between the external pressure and the vapor pressure of water from the solution. By varying the external pressure, the partial pressure p of
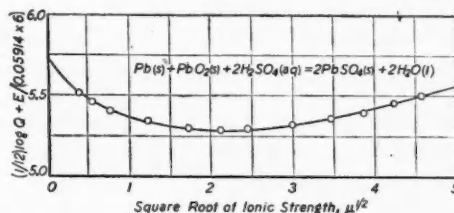
FIG. 14. The mean activity coefficient of sulfuric acid and the standard electromotive force in the lead-acid storage cell. When $E/(6 \times 0.05914) + (1/12) \log Q$ is plotted against $\mu^{\frac{1}{2}}$, the value of the standard electrode potential is obtained from the intercept on the axis of $\mu^{\frac{1}{2}} = 0$, in accordance with Eq. (40). Fig. 14 is drawn on a different scale from Fig. 9, but the limiting slope is the same.

TABLE IV. *Variation of voltage of lead-acid storage cell with specific gravity in limited range at 25°C.*[24]

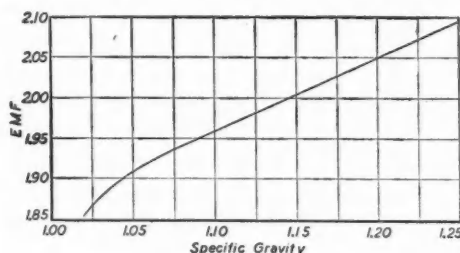| Sp. gr.	1.02	1.03	1.04	1.05	1.10	1.15	1.25
Emf.	1.854	1.878	1.894	1.907	1.961	2.005	2.098

FIG. 15. The reversible potential of the lead-acid storage cell at 25°C. The specific gravities plotted here are those of Table IV. In the storage battery operating under normal conditions the electromotive forces for the several specific gravities would differ slightly from those shown on the curve because the density of the acid in direct contact with the active material would not be exactly the same as that in the body of the electrolyte.

the hydrogen is also changed. If this is the only variable we may write

$$E = const. + (RT/2F) \ln p. \qquad (42)$$

This equation was first tested by Lewis[25] who allowed the hydrogen in such a cell to escape against a variable excess pressure. The results (at 25°C) are given in Table V. The excess pressure is given in centimeters of water. (Between the fourth and fifth experiments the excess pressure was raised to 100 cm.)

The very satisfactory agreement between the observed and the calculated values, together with the lack of appreciable polarization in the electrode, make it certain that we are dealing here with a well-defined and reversible reaction. Some years ago a similar oxygen electrode was often employed, but has since been shown by indirect methods to give quite erroneous results. This would have been detected earlier if similar experiments had been made with varying partial pressures of oxygen.[26]

[25] Lewis and Randall, J. Am. Chem. Soc. **36**, 1969 (1914).

[26] The interesting case, studied by Edgar, of the sulfur dioxide electrode in sulfuric acid, which leads to erroneous results in spite of the fact that the change of the electromotive force with the pressure is approximately the one predicted, is discussed by Lewis, Randall and Bichowsky, J. Am. Chem. Soc. **40**, 356 (1918).

VARIATION OF THE ELECTROMOTIVE FORCE WITH TEMPERATURE

Still another check upon the correctness of an assumed cell process is furnished by the variation of the electromotive force with the temperature. The fundamental equation for the change of electromotive force with temperature (known as the Gibbs-Helmholtz equation) is

$$nF(dE/dT) = -\Delta S = (-nFE - \Delta H)/T. \qquad (43)$$

Very extensive studies of the temperature coefficient of the electromotive force of galvanic cells were made by Gerke.[27] His values for the increase in heat content in the reactions corresponding to his measured cells agree well with the measured calorimetric values, thus demonstrating that the processes which he assumed to take place in his cells were probably correctly formulated. Lewis[28] carefully measured the electromotive force of the hydrogen-calomel cell at various temperatures. He calculated values not only of the increase in heat content but also of the change of the increase of heat content with the temperature corresponding to the reaction of Eq. (30) and found that these quantities corresponded satisfactorily with the measured calorimetric values for the increase in heat content and heat capacity, respectively. The determination of the coefficient of the electromotive force of a cell and its interpretation by means of the Gibbs-Helmholtz equation furnish one of the satisfactory means of determining calorimetric quantities, provided it is possible to obtain thoroughly satisfactory reversible and nonpolarizable electrodes.

ELECTROLYTIC CELLS DISCHARGING AND CHARGING AT A FINITE RATE

The efficiency of an actual machine is always less than unity. There is always some loss due to

TABLE V. *Effect of pressure on the hydrogen electrode*

Δp (CM OF H_2O)	0	37	63	84	84	63	37
E (Obs.)	0.40089	0.40134	0.40163	0.40190	0.40189	0.40164	0.40138
E (Calc.)	.40089	.40134	.40165	.40189	.40189	.40165	.40134

[27] Gerke, J. Am. Chem. Soc. **44**, 1684 (1922).

[28] See Ref. 26. These measurements were later repeated by Ellis [J. Am. Chem. Soc. **38**, 737 (1916)] but his results proved to be somewhat inferior in accuracy. The acid in the cell was not precisely 0.1 M, but was sufficiently close to it for our present purpose. Another method of utilizing these data is given in Ref. 26. The present method is, however, simpler and has a wider range of validity.

friction, regardless of how perfectly the machine may be constructed; for instance, the mechanical energy which may be recovered from a flywheel is always less than that used to set the flywheel in motion. While analogies are sometimes misleading, it is well to consider the difference in the charging and discharging potentials of a storage cell or other cell as due to the sum of several effects, each of which functions in the same way as does friction in a mechanical system. As friction in a machine increases with increasing pressure on the bearing, so does "friction" in the electrical system increase as a larger current is taken from the cell.

FIG. 16. Typical voltage and gravity characteristics of a lead-acid storage battery. These characteristics will vary slightly for cells of various designs owing to varying rates of diffusion, stirring, internal resistance and proportion of acid in the pores of the plates. [Courtesy Electric Storage Battery Co.]

While a storage cell is discharging, the voltage at the terminals is always less than the reversible electromotive force and falls off as the discharge proceeds. On the other hand, the charging voltage is always larger than the reversible electromotive force and rises as the cell is charged.

Classification of losses in an operating cell.— The *total* or *energy, efficiency* of an electrolytic cell is the product of the voltage efficiency and the ampere-hour efficiency. The *ampere-hour efficiency* (so-called "current efficiency") is the ratio of the mols of electron obtained to the equivalents of material actually used in the reactions of charge and discharge. The loss of voltage efficiency depends mainly upon the losses due to the resistance of the cell and to the slowness of diffusion of the solutions held within the fine pores of the plates.

Ampere-hour efficiency.—The ampere-hour efficiency of a secondary cell is dependent upon the ratio of quantity of electricity recovered in the discharge-half of the cycle to that used in charging the cell back to the condition existing at the beginning of the previous discharge half-cycle. It is an index of the extent to which the construction and condition of a cell insures the absence of side-reactions. Losses in ampere-hour

efficiency are analogous to leakages due to corona losses from a high voltage transmission line or to leakage of water at badly made joints in a water distribution system. As an example of the influence of side-reactions upon ampere-hour efficiency, we may illustrate by a discussion of some of the side-reactions in the lead-acid storage cell.

The positive plate of the cell consists of a porous mass of lead peroxide held within a conducting grid of antimonial lead. In the course of discharge, this lead peroxide reacts with sulfuric acid to form lead sulfate and water in accordance with Eq. (5). Now on recharging of the cell, lead sulfate and water are reconverted into lead peroxide and sulfuric acid. Lead sulfate and lead peroxide are not very good conductors of electricity. Colloidal lead sulfate which is produced by the discharge-reaction in contact with sulfuric acid tends to crystallize and form a crystalline variety of lead sulfate. Obviously the amount of crystalline lead sulfate which is formed in the plate will be roughly a product of the amount of lead sulfate present and its time of contact with the sulfuric acid. A similar phenomenon takes place in the negative plate.

On recharging the battery only a portion of this crystalline lead sulfate is reconverted to lead peroxide and, hence, we have the phenomenon

of "sulfating" of the positive and negative plates. To the extent then to which "crystalline" lead sulfate is formed during each cycle, we have a successive lowering in the ampere-hour capacity of the cell.

A side-reaction causing a lowering of the ampere-hour capacity and efficiency of the storage cell is the direct formation of lead sulfate by chemical action of sulfuric acid upon the lead peroxide to form lead sulfate, water and oxygen. This is a very slow reaction. A similar reaction is the action of sulfuric acid upon the lead of the cathode to form lead sulfate and hydrogen. These reactions will not ordinarily occur unless there is some impurity or "catalyst" that will favor such a reaction.

The "internal short-circuiting" of the cell resulting from a particle, impurity or catalyst that allows the main reaction of discharge to proceed without our obtaining useful work at the terminals is a true side-reaction. The grid of the positive plate is antimonial lead, which may act very slowly as a short-circuited negative plate. This action ultimately results, after many cycles, in the gradual conversion of the positive grid into active material.

Another cause of ampere-hour inefficiency is the side-reaction in which the positive and negative plates of a cell act as inert electrodes and the charging current serves to electrolyze water with a discharge of oxygen and hydrogen at the plates. Such inefficiency is largest when the cell is almost fully charged or when it is fully charged and the current through the cell is continued (Overcharge). The discharge of gaseous hydrogen and oxygen at the plates is, of course, also destructive to the plates as it tends to loosen fine particles of active material. This gassing is the cause of the greatest wear upon the plates. The loosened particles may fall to the bottom of the battery as sediment or may possibly be carried by circulation of the electrolyte to the opposite plate. They may cause internal short-circuiting of the plate with further loss of efficiency. Of course, the continued charging of the cell tends to ultimately convert partially crystallized lead sulfate into lead peroxide and lead, thus restoring somewhat the ampere-hour capacity per cycle. Another possible side-reaction at very high rates of charge is the

electrolysis of water at the lead grids, instead of the main reaction of Eq. (5).

Voltage efficiency.—Ordinarily, the larger the rate of charge and discharge of the cell, the larger the loss of voltage efficiency. We may separate the various potentials making up the measured potential of any actual cell into the following classes: reversible potential, E_{rev}; internal resistance loss, E_{res}; concentration polarization, E_{conc}; and polarization, E_{pol}. If E is the potential across the terminals then, on discharging a cell,

$$E = E_{rev} - E_{res} - E_{conc} - E_{pol}, \qquad (44)$$

and on charging a cell, or in electrolysis,

$$E = E_{rev} + E_{res} + E_{conc} + E_{pol}. \qquad (45)$$

The reversible potential has already been fully discussed in previous sections. We found that the reversible potential is given experimentally when the electrodes are "active" and of ample size, and the current is so low that the charging and discharging potentials of the cell are equal.

Losses due to internal resistance.—The total resistance of a cell is made up of the resistances of the metallic electrodes and of the electrolyte. In addition, there may be an internal resistance due to the presence of a film of poorly conducting gas, liquid or solid on one or both of the electrodes. Since the effect of a film is usually considered in connection with polarization, we shall assume in the discussion of E_{res} that the reactions at the electrodes are truly reversible. The potential drop across the cell due to the internal resistance is zero when there is no current through the cell, this being one of the conditions for equilibrium. It increases directly with the current. By Ohm's law, E_{res} in Eqs. (44) and (45) is the quotient of the current and the total resistance of the cell as here defined. Since the resistance of a solution decreases with increase in temperature, this E_{res} becomes smaller as the temperature of the electrolyte rises.

The internal resistance of a cell depends upon the shape of a cell or, rather, upon the shape of the electrolyte lying between the plates of the cell. Obviously, the larger the plates, the greater will be the cross-sectional area of the electrolyte and, hence, the smaller the resistance. On the other hand, the smaller the distance between the

plates, the smaller will be the resistance and, hence, the smaller E_{res}.

To make sure that there will be no physical contact between the plates of a cell, and to avoid "short-circuiting" between the electrodes or plates, separators are often used. Obviously, any separator will increase the resistance of a cell owing to the interposition of nonconducting materials. If the separator is made of very porous wood, of very thin perforated hard rubber or of very loosely woven glass fiber cloth or rubber inserted cloth, the increase in resistance will not be very large.

The resistance of the electrolyte also varies with the concentration, and the value of E_{res} can be reduced by choosing electrolytes having a maximum conductivity. In some cases, as with cells used for electroplating, or for the electrolytic winning or refining of metals, it is the practice to add some electrolyte of high conductivity in order to decrease the resistance of the cell. Thus, in the case of the copper refining cell, the electrolyte consists largely of sulfuric acid which has a much higher conductivity and is much more soluble than copper sulfate.

The mechanism of electrolysis.—In order to understand the inefficiencies relating to concentration polarization we must understand the mechanism of conduction in electrolytes. For simplicity let us consider two silver electrodes in a cell such as that shown diagrammatically in Fig. 17 with aqueous silver nitrate as the electrolyte. We write the half-reactions for this cell as follows:

$Ag = Ag^+ + E^-$; $\mathbf{E}^0 = -0.7995v$;
$E = -0.7995 - 0.05914$
$\qquad \times \log a(Ag^+)$;   (46)

$Ag^+ + E^- = Ag$; $\mathbf{E}^0 = 0.7995v$;
$E = 0.7995 - 0.5914$
$\qquad \times \log [1/a(Ag^+)]$.   (47)

It is obvious that, if the concentration of the silver ion is the same at the two electrodes, the electromotive force obtained by adding the potentials of the two half-reactions is zero.

Let us assume that the anode and cathode compartments of

the cell each contains 10 l of molal silver nitrate solution, and that for the moment there is no transfer of silver ions or of nitrate ions across the boundary between these compartments. Let us now pass 1 mol, or 1 faraday, of electricity through the cell. In the anode compartment 1 mol of silver will be converted into silver ion in accordance with Eq. (46). At the same time in the cathode compartment 1 mol of silver ion will be changed to silver in accordance with Eq. (47). If there were no movement of ions across the boundary, the average concentration of silver ion in the anode compartment would now be 1.1M, while that in the cathode would be 0.9M. The average concentration of the nitrate ion would remain 1M throughout the cell. However, this condition cannot actually exist, for the solutions necessarily maintain a condition of electroneutrality.[7] What then is the solution of this riddle?

Metallic and electrolytic conductors.—A metallic conductor contains metallic ions and more or less firmly bound "free" electrons. The relative mobility of the metallic ions and the electrons is so greatly in favor of the latter that, when a metal conductor is placed under the influence of a potential gradient, there is no noticeable transport of material other than electron. An electrolytic conductor, on the other hand, contains both positive and negative ions with no free electrons. If, therefore, an electrolytic conductor is placed in a potential gradient, the positively charged particles will move through the electro-
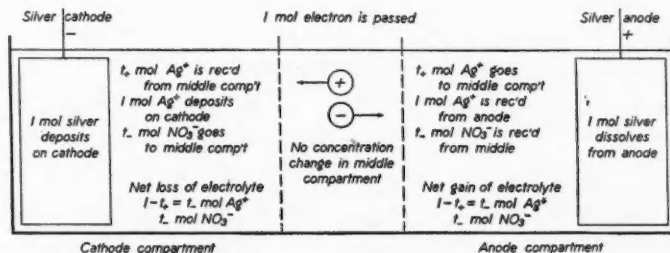
FIG. 17. Hittorf's transference experiment. His classical apparatus [Ann. d. physik. Chemie 89, 177 (1853); 98, 1 (1856); 103, 1 (1858); 106, 337 (1859)] may be diagrammatically represented as a trough divided into 3 compartments by porous partitions which prevent mechanical mixing of the several portions of the electrolyte, but which do not hinder the free movement of the ions under the influence of a potential gradient. If the concentration of electrolyte in the middle compartment is the same at the beginning and at the end of an experiment, we may assume that all of the concentration changes were due to electrical and chemical effects.

lyte toward the cathode and the negatively charged particles will move in the opposite direction toward the anode.

If we make a cell with large parallel electrodes 1 cm apart, with 1 v difference of potential, the current per square centimeter (that is, the conductance of a column of solution 1 cm² in cross-sectional area when the electrodes are so large that the electrolyte may be regarded as a slab of infinite area and 1 cm thickness) is defined as the specific conductance λ. If λ_c is a specific conductance and c is the concentration of the electrolyte in mols per liter, the volume of electrolyte contained between the electrodes 1 cm apart which contains 1 mol of electrolyte is $1000/c$. Hence the number of unit cubes is $1000/c$, each of a conductance λ_c, and the *mol conductance* Λ_c, or conductance per mol, is

$$\Lambda_c = 1000\lambda_c/c. \qquad (48)$$

If the potential between the electrodes of the foregoing cell which contains 1 mol of electrolyte is 1 v, then 1 mol of electricity (96,489 coulombs) will be carried per second per equivalent of ion moving from one electrode to the other at the speed of 1 cm/sec.

On the basis of our definition of the molal conductance of a salt, we may define the mol conductance of an ion as the quantity of electricity carried per second past a hypothetical plane or cross section of a solution by 1 mol of that ion
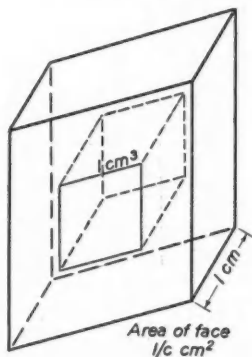
Fig. 18. Relation between specific and molal conductance. Assume a conductivity cell with ends consisting of parallel electrodes 1 cm apart. Assume the area of the electrodes to be such that a volume of solution containing 1 mol of the electrolyte is present. If c is the concentration of the electrolyte in mols per liter, the volume of the electrolyte and, hence, the number of unit cubes, is $1000/c$.

under a potential gradient of 1 v/cm. One mol of a positive ion of valence z_+ will therefore carry 96,489 z_+ coulombs of electricity. Hence the molal conductance of an ion is 96,489 z_+u_+, where u_+ is the absolute speed or mobility of the ion in centimeters per second under a potential gradient of 1 v/cm. Because of the necessity of electroneutrality in the solution and the simultaneous movement of positive and negative ions, the molal conductivity of a cell is made up of the mols of electricity carried by the positive ions plus the mols of electricity carried by the negative ions.

The proportion of the electricity carried by the positive ion to that carried by the negative ion is seen to be proportional to the absolute speeds of the ions. The proportion carried by the positive ion is therefore called the *transference number* of the positive ion. It is defined by the equation

$$t_+ = u_+z_+/(u_+z_+ + u_-z_-), \qquad (49)$$

where u_+ and u_- are the respective speeds of the ions under a potential gradient of 1 v/cm.

It is found by experiment that the silver ions in molal silver nitrate solution travel toward the cathode at the rate of 0.00057 cm/sec under a potential gradient of 1 v/cm at 18°C and that the nitrate ions travel toward the anode at the rate of 0.00064 cm/sec under the same potential gradient. In other words, $t_+ = 0.00057/(0.00057 + 0.00064) = 0.471$ is the fraction of the current carried by the silver ions and $t_- = 0.00064/0.00121 = 0.529$ is carried by nitrate ions. Let us now look at the result in our electrolysis cell after the passage of 1 faraday of electricity.

Concentration changes during electrolysis.—In the anode compartment 1 mol of silver ion has been formed, and 0.471 mol of silver ion has moved out of the anode into the cathode compartment. At the same time 0.529 mol of nitrate ion has been carried from the cathode to the anode compartment. We now find the concentration of the silver ion to be 1.0529M and that of the nitrate ion to be 1.0529M, or the solution in the anode compartment is electrically neutral. By a similar calculation we find that the concentration of both silver and nitrate ions in the cathode compartment is now 0.947M, and the

solution is again electrically neutral. The net result of our electrolysis is the disappearance of a mol of silver at one electrode and its appearance at the other.

Since the activity of the metallic silver in each case is unity, no work has been performed during this transfer of metallic silver. On the other hand, the silver nitrate solution, originally at a uniform concentration, has been altered so that part of it is at a lower, and part at a higher, concentration than at the beginning of the electrolysis. It is obvious that this process involves the expenditure of work and an increase in the free energy of the system, since any solution at a non-uniform concentration tends to proceed spontaneously to a condition of uniform concentration. Stated in other words, a uniform solution left to itself never alters spontaneously to produce regions of varying concentration.

The potential of the concentration polarization is therefore measured by the work required to transport the more dilute electrolyte, in this case from the immediate vicinity of the cathode to the higher concentration in the immediate vicinity of the anode.[29] The larger the difference in concentration, the larger is this potential. The concentration difference in turn depends upon the current density, the size and shape of the solution and upon whether or not the solution is stirred. If it is not stirred, the difference depends also upon the rate of natural convection and diffusion. As in the case of the loss due to resistance, an increase in temperature increases the rate of diffusion and convection, thus causing a smaller concentration polarization for a given current density.

The conductance of a solution that has uniform concentration is always larger than the conductance of one containing the same total amount of electrolyte and the same volume and shape but of non-uniform concentration. The segregation of the electrolyte tends to increase the internal resistance of the cell and therefore increases the $\mathbf{E}_{res}$ loss.

The potential of concentration cells with transference.—We find the concentration of the sulfuric acid in the vicinity of both the anode and cathode of the lead-acid storage cell decreasing

because sulfuric acid is used up in the discharge reaction [Eq. (6)]. However, owing to the transference of sulfuric acid in accordance with the principle discussed in the preceding paragraphs, we find that the concentration of sulfuric acid is decreased more at the anode than at the cathode because sulfuric acid has been transported out of the anode compartment during the discharge. Thus, a portion of the decrease of the potential of the lead-acid storage cell is due to a change in the activity of the sulfuric acid in accordance with Eq. (10).

In considering the electromotive force due to the difference in concentration at the electrodes we might think of a cell consisting of an anode of sulfuric acid at concentration x_aM with lead peroxide and lead, and the other electrode of sulfuric acid (x_cM) and lead peroxide and lead electrodes as indicated schematically by the equation

$$Pb(s) \,|\, PbO_2(s) \,|\, H_2SO_4 \; x_aM \,|\,| \\ H_2SO_4 \; x_cM \,|\, PbO_2(s) \,|\, Pb(s) \quad (50)$$
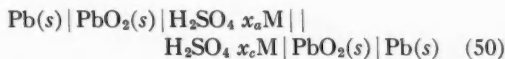
Such a cell with reversible electrodes will give a definite and reproducible electromotive force.

Nernst,[30] in the early days of the ionic theory, showed how such a cell could be treated thermodynamically. In his solution of the problem he used ion concentrations (obtained from conductivity measurements), where now we should use the activities, and his equation gave results that were only approximately correct. The problem is, however, one with which thermodynamics is perfectly competent to deal in an exact manner.

In solving this problem we may, for the sake of variety, make use of the partial molal free energies instead of the activities. In order to fix our ideas, let us consider the concentration cell as sketched in Fig. 19. At A and to the left of A the molality of the acid is constant and equal to m^A;
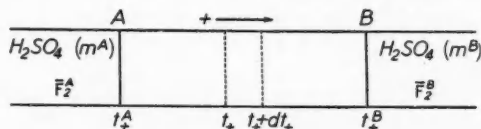
FIG. 19. The electromotive force of a cell with transference is equivalent to the work done (in proper units) when the electrolyte is carried reversibly from the one concentration to the other. In the diagram 1 mol of sulfuric acid is carried from molality A to molality B.

[29] The potential of such cells with transference is discussed at length in Ref. 7(a).

[30] Nernst, Zeits. f. physik. Chemie 4, 129 (1889).

the transference number of the cation is t_+^A, and the partial molal free energy of the $H_2SO_4(aq)$ is $\bar{F}^A$. Somewhere between A and B the concentration varies in some continuous but unspecified manner, until at B, and to the right, the molality is constant at m^B, the transference number at t_+^B and the partial molal free energy at $\bar{F}^B$.

Now when the cell operates, and a small negative current passes through the cell from right to left, the total change in free energy per 2 equivalents of electricity (1 mol of sulfuric acid contains 2 equivalents) is determined if we know the amount of acid being added or removed at each point and also the value of the partial molal free energy at each point.

The amount of acid leaving the region to the left of A is t_+^A, the amount entering the region to the right of B is t_+^B and the difference is taken from (or given to) the intervening region where the gradient occurs. In the region to the right of B the free energy increases by $t_+^B \bar{F}^B$; in the region to the left of A it increases by $-t_+^A \bar{F}^A$. In the intervening region there is a summation of effects which may be analyzed as follows: Consider some infinitesimal region (indicated by the space between the dotted lines) in which the transference number varies from t_+ on the one side to $t_+ + dt_+$ on the other, and in which the partial molal free energy is $\bar{F}$. The amount of acid leaving this region is dt_+, hence the increase in free energy in the total intermediate region is $-\int_A^B \bar{F} dt_+$. Therefore, for the whole cell,[21]

$$\Delta F = t_+^B \bar{F}^B - t_+^A \bar{F}^A - \int_A^B \bar{F} dt = \int_A^B t_+ d\bar{F}. \qquad (51)$$

Now, since $n = 2$, we have

$$\Delta F = -2\mathbf{E}F = -2 \times 23\,054\mathbf{E} \qquad (54)$$

and

$$\mathbf{E} = -(1/2\mathbf{F}) \int_A^B t_+ d\bar{F} = -(1/2\mathbf{F}) \int_A^B t_+ d(\bar{F} - \bar{F}^0).$$

We may therefore calculate the electromotive force of any such concentration cell if we know, over the range of concentration, the values of the transference numbers and of the partial molal free energies (which may, for example, be obtained from the cells without liquid junctions). It is necessary only to plot at several concentrations the values of t_+ against the values of $\bar{F} - \bar{F}^0$, and evaluate the area under the curve, divided by $-2\mathbf{F}$, which gives immediately the desired result.

Substituting $RT \ln a$ for $\bar{F} - \bar{F}^0$ we obtain

$$\mathbf{E} = -(RT/2\mathbf{F}) \int_A^B t_+ d \ln a$$
$$= -(3RT/2\mathbf{F}) \int_A^B t_+ d \ln m_+ \gamma_+. \qquad (54)$$

Eq. (54) merely expresses in terms of electromotive force the work necessary to carry 1 equivalent of sulfuric acid in the dilute concentration in which it exists at the anode to the more concentrated condition in which it exists at the cathode. This work is the true concentration polarization of the cell and is, of course, diminished to zero for a cell in which the electrolytes are stirred. For static electrolytes the value of the concentration polarization will depend

[21] $\int x dy = xy - \int y dx.$ (53)

upon the rate of change of the concentration of the electrolyte due to the rate of discharge of the cell and upon the effect of diffusion in neutralizing or destroying this difference in concentration. The lead-acid storage cell is purposely built so that the plates are very porous, and hence the development of this concentration polarization is quite rapid because it is not easy for the diffusion process to neutralize the effect of the concentration polarization.

Overvoltage.—The term $\mathbf{E}_{pol}$ in Eq. (44) relates to various types of voltage inefficiencies called *polarization* or *overvoltage*. If a rod of amalgamated zinc is immersed in a mixture of molal hydrochloric acid and molal zinc chloride, there is no visible evolution of hydrogen. However, if we consider the reaction,

$$Zn + 2H^+ = Zn^{++} + H_2(g);$$
$$\mathbf{E}^0 = 0.758 \text{ v, } \Delta F^0 = -34,980 \text{ cal,} \qquad (55)$$
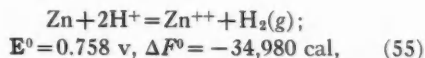
it is seen that the pressure of the hydrogen at equilibrium for molal zinc ion is 4×10^{25} atmos. [Eq. (10).] In other words, the reaction should proceed spontaneously against hydrogen at a pressure of 1 atmos. However, only a very slight evolution of hydrogen will take place at the amalgamated zinc surface until a potential of 0.60 v is applied. Table VI shows the *overvoltage* values for hydrogen from sulfuric acid solution at several different surfaces. It is seen that the overvoltage depends on the physical state or character as well as upon the chemical composition of the surface. The theory for the cause of overvoltage is somewhat in dispute. Some attribute it to concentration polarization, limited to the film adjacent to the electrode; others

TABLE VI. *Cathodic overvoltages for hydrogen in molal sulfuric acid solution at 25°C (International Critical Tables).*

NATURE OF CATHODE	CURRENT DENSITY (AMP/CM²)		
	0.0001	0.01	0.1
	OVERVOLTAGE (v)		
Smooth platinum		0.068	0.29
Platinized platinum	0.0034	0.030	0.041
Gold	.122	0.390	0.59
Mercury	.60	1.04	1.07
Copper	.351	0.584	0.80
Tin	.400	1.08	1.22
Iron	.218	0.56	0.82
Zinc		0.75	1.06
Cadmium	.651	1.134	1.216
Graphite	.317	0.779	0.977

consider it due to the slowness of the reaction at the electrode, etc.

It is seen that during the discharge of the lead-acid storage cell all known irreversible effects, as well as the dilution of the sulfuric acid, tend to decrease the voltage. We might define the voltage efficiency of the discharge of the cell as the measured voltage divided by the reversible voltage. Fig. 16 shows the measured voltage at several rates of charge and discharge for the lead-acid storage cell. The irreversible voltages all tend to have a larger effect the larger the rate, and this is verified experimentally, as shown by the curves of Fig. 20.

CONVENTION FOR THE SIGN OF ELECTROMOTIVE FORCE

When the cell is represented by the equation for the process which is taking place in the cell, the convention regarding the sign of the electromotive force of a cell is in agreement with the universal custom of chemists. The reaction is written in such a way that the substances which disappear occur in the left-hand member of the equation and the substances which appear occur in the right-hand member. Then the sign of the electromotive force will be positive when the reaction is spontaneous and will be negative when work must be supplied to cause the reaction to proceed.[8]

It is not always convenient, especially before the nature of the chemical reaction has been definitely established, to write the equation for the process taking place in the cell. We may then express the cell diagrammatically as was done in Eq. (50). For example, the lead-acid storage cell may be written

$$PbO_2(s), \ PbSO_4(s), \ H_2SO_4(xM),$$
$$PbSO_4(s), \ Pb(s). \quad (56)$$
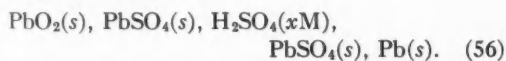
We shall adopt the convention that a positive electromotive force represents a positive tendency of the negative current to pass spontaneously through the cell from *right to left*. Thus at 25°C we write $\mathbf{E} = 2.1$. If we represent the cell diagrammatically as

$$Pb(s), \ PbSO_4(s), \ H_2SO_4(xM),$$
$$PbSO_4(s), \ PbO_2(s), \quad (57)$$
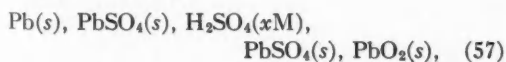
then we must, in accordance with our convention,
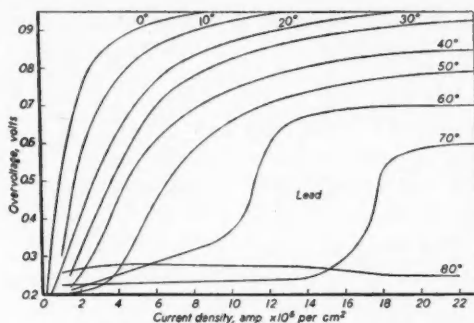
FIG. 20. Relation of the overvoltage of lead to the current density at several temperatures. At the lowest temperature the overvoltage of the lead rises rapidly with the current density, but as the temperature of the 0.05M aqueous sulfuric acid is increased, the overvoltage at a given current density and the rate of increase with current density both are reduced. We may note that the fluidity, rate of convection, absolute speed and rate of diffusion of the ions all are increased by increased temperature. [Bircher, Harkins and Dietrichson, J. Am. Chem. Soc. 46, 2622 (1924).]

write $\mathbf{E} = -2.1$. Expression (57) is a conventional method of stating in reverse form the same information as is given in Expression (56).

Of course, our knowledge of the chemistry of the cell usually prevents an error in sign, even though no conventions such as the foregoing are observed; but this is by no means always the case. Thus two electrodes, one of silver and silver chloride, the other of mercury and mercurous chloride, both in the same solution of some chloride, yield an electromotive force of 0.0455 v. Not all chemists could state offhand whether the silver or the mercury is consumed when the cell operates spontaneously. However, if we write

$$Ag \ AgCl, \ KCl(aq), \ HgCl, \ Hg;$$
$$\mathbf{E}_{298.1} = 0.0455, \quad (58)$$
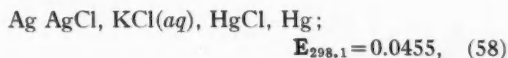
we are at once informed not only of the magnitude of the electromotive force but also that silver spontaneously replaces mercury in the chloride, since this is the process which accompanies the passage of a negative current through the cell from right to left.

By the similar and entirely compatible convention that we also use, we merely write

$$Ag(s) + HgCl(s) = AgCl(s) + Hg(l);$$
$$\mathbf{E}_{298.1} = 0.0455, \quad (59)$$
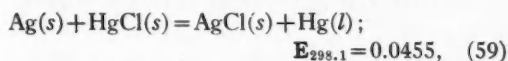
for a positive value of $\mathbf{E}$ always means that the reaction as written proceeds spontaneously so

that the substances on the left are consumed and those on the right are produced.

If we are to use the two methods simultaneously, in other words, if we are to state both the composition of the cell and the cell reaction, then, in order to prevent inconsistency in our conventions, we must write the reaction in the way in which it will occur when the negative current passes from right to left through the cell as specified. For example,

$$\text{Ag, AgCl, ZnCl}_2(aq), \text{Zn}; \mathbf{E} = -;$$
$$2\text{Ag} + \text{ZnCl}_2(aq) = 2\text{AgCl} + \text{Zn}. \quad \mathbf{E} = -. \quad (60)$$

In accordance with these conventions we may name the electrodes of a cell in the following manner. The *anode* is that electrode at which reaction of the material in the half-cell liberates electrons; for example, the lead-acid storage cell discharging in accordance with Eq. (4). This electrode is ordinarily called the *negative electrode* by physicists and engineers.

For the same reason the lead peroxide electrode which is reduced in accordance with Eq. (5) is called the *cathode* or the electrode at which electrons are used up. In accordance with the convention of the physicist and engineer, the lead peroxide electrode is known as the *positive electrode*.

Electrical instruments are so calibrated that, when the plus terminal is attached to the cathode of a cell spontaneously discharging, and the other to the anode, the instrument will show a positive reading.

While there can be no ambiguity as to what is meant by positive electromotive force when the reaction of the cell is written as in Eqs. (4), (5) and (6), there is also no ambiguity as to which electrode is positive and which electrode is negative, and by the convention of Eq. (56) there is no ambiguity in stating that the electromotive force is positive when the negative electricity spontaneously flows from right to left within the cell as diagrammed. But the question as to whether or not we shall write $\text{PbO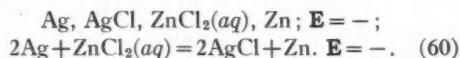}_2$ or PbSO_4 first in the diagrammatic representation of the process at the anode is not always clear. It is customary to write the lead peroxide first but not because it disappears. The lead peroxide and lead sulfate are equally necessary to the establishment of a reversible electromotive force at the anode. Likewise, it is customary to write $\text{Pb}(s)$ last at the cathode of the discharging lead-acid storage cell, probably because the lead not only is a part of the process but also serves to conduct the negative current into the cell.

Exact Averages for Use in the Teaching of Elementary Physics

PAUL L. COPELAND
Department of Physics, Armour Institute of Technology, Chicago, Illinois

TEACHERS of elementary physics frequently have need to justify a result that is ordinarily derived by the use of the integral calculus. Various purely algebraic methods are available for this purpose. One of these has been presented by Bacon,[1] his method being to obtain algebraically the sum of a number of terms which is allowed to approach infinity as a limit. This is a straightforward process which has the advantage of introducing some of the important ideas of the integral calculus without confusing the student with an unfamiliar notation. There is in the literature another presentation which some

teachers may prefer, because formally it is more simple. In it the process of passing to a limit is not necessary, although it can be used if desired, and the ignoring of small terms, which vanish in the limit, is entirely eliminated.

A very simple illustration of this method is given by Duncan and Starling[2] in the process of determining the potential difference between two points A and B due to a point charge Q. The work involved in moving a unit charge along a line extending radially outward from the point charge is the average electrostatic field multiplied

[1] R. H. Bacon, "Moments of Inertia for First-Year Physics Students," Am. Phys. Teacher **6**, 72 (1938).

[2] Duncan and Starling, *A Textbook of Physics* (Macmillan, 1922), pp. 930–31. The same illustration appears in Spinney, *A Textbook of Physics* (Macmillan, 1937).

by the distance between the initial and final positions. Duncan and Starling break the distance between the two points A and B into a number of short steps A to X, X to Y, Y to Z, and so on, ending with a step from K to B. If a represents the distance from A to the point charge Q and if x represents the distance from X to the point charge, we see that the field varies in the first step from a value of Q/a^2 at the beginning to a value of Q/x^2 at the end. The average value will lie somewhere between them, and Duncan and Starling introduce the entirely reasonable assumption that the average value of the field between A and X may be represented by Q/ax. Attention is not called to the fact that this is the *exact value* of the *average*, however, and yet it is this fact that simplifies the subsequent argument. When the average field Q/ax is multiplied by the distance $x-a$, we find the work done by the charge in the interval to be $Q(x-a)/ax$ or $(Q/a)-(Q/x)$. Similarly, between points X and Y, the work done is $(Q/x)-(Q/y)$. It thus turns out that in the final summation all of the intermediate terms drop out by pairs, leaving, for the final answer, the familiar result $(Q/a)-(Q/b)$. In this argument the necessity of proceeding to a limit to eliminate small terms is avoided by the use of the exact value of the average.

The crux of the problem involved in generalizing the procedure illustrated by Duncan and Starling is to determine the average of x^n for a definite interval in terms of the values of x at the limits. This can be done by a simple algebraic procedure. If n is a positive integer, such an expression may be obtained by averaging x^n over a set of $(n+1)$ values of x in geometric progression.[3, 4] If the interval over which we wish to obtain an average of x^n begins at a and ends at b, we let $r^n=b/a$. The values of x for which we average x^n are then a, ar, $ar^2\cdots ar^n$, and the average is

$$\{1/(n+1)\}\{a^n+(ar)^n+(ar^2)^n+\cdots+(ar^n)^n\}.$$

But, since $ar^n=b$, we have at once for the average of x^n between $x=a$ and $x=b$:

[3] The writer is indebted to Professor L. R. Ford of the Department of Mathematics for suggesting this method of arriving at the average.

[4] This method of averaging is used by Edser throughout his *General Physics for Students* (Macmillan, 1913), Chap. 2, especially pp. 48–49.

$$\overline{x_{ab}{}^n}=\{1/(n+1)\}$$
$$\times\{a^n+a^{n-1}b+\cdots+ab^{n-1}+b^n\}. \quad (1)$$

For a negative power, x^{-n}, we have only to average over the $(n-1)$ interior points of the same set; thus,

$$\overline{x_{ab}{}^{-n}}=\{1/(n-1)\}\{1/(ar)^n+1/(ar^2)^n$$
$$+\cdots+1/(ar^{n-1})^n\}$$
$$=\{1/(n-1)\}\{1/a^{n-1}b+1/a^{n-2}b^2$$
$$+\cdots+1/ab^{n-1}\}. \quad (2)$$

In connection with Eq. (2) it may be noticed that if $n=2$, the average is $1/ab$, which is the result used by Duncan and Starling in the example cited. Also, it may be noticed that if n is set equal to unity, the expression becomes indeterminate; this is the single case where n is an integer for which the average is not a simple algebraic expression. In connection with Eq. (1) it should be noticed that if $n=1$, the expression for the average reduces to $(b+a)/2$, an expression with which the student is familiar and which he has used to obtain the average in cases of linear variation. It should be emphasized, however, that the average is the arithmetical mean of the limits only if $n=1$. Geometrical relationships, with which the students are familiar, may be used to illustrate the application to cases for which $n=2$. For example, the volume of a cone or a pyramid, or of a truncated cone or pyramid, may be found by multiplying the average cross-sectional area by the height. Such illustrations do much to show where the cases with which the student is familiar fit into the general picture.

When the average value of x^n is multiplied by the length of the interval, $b-a$, the result is x^n summed throughout the interval. Thus for positive powers we have

$$\overline{x_{ab}{}^n}(b-a)=(b^{n+1}-a^{n+1})/(n+1) \quad (3)$$

and, for negative powers,

$$\overline{x_{ab}{}^{-n}}(b-a)=(b^{-n+1}-a^{-n+1})/(-n+1). \quad (4)$$

The right-hand members of these expressions will be recognized at once as results ordinarily obtained by integration. Furthermore, these expressions for the sum are valid even for fractional exponents. The average value for fractional exponents may be obtained from these expressions directly by long division. They are not polonomials but infinite series.

The area under a curve may be obtained at once from the expression for the average. We illustrate this through a general determination of the area lying under the curve x^n between $x=a$ and $x=b$. For completeness at this point we

divide the interval between a and b into steps a to c, c to d, d to e, and so on, ending with w to b, as shown in Fig. 1. The value of x^n varies in the first step from a^n at the beginning to c^n at the end. Its average lies somewhere between these limits and we represent it by the expression obtained previously, knowing that if this expression is not exact, we can get a result to any desired accuracy by the expedient of making the steps small. The average in the first step, $(a^n+a^{n-1}c+\cdots+ac^{n-1}+c^n)/(n+1)$, when multiplied by the length of the step, $c-a$, yields $(c^{n+1}-a^{n+1})/(n+1)$ at the contribution to the area. Similarly in the next step, using an expression of the same form for the average of x^n within this step, we obtain $(d^{n+1}-c^{n+1})/(n+1)$ as the contribution to the area. Similar expressions represent the contributions to the area in the various steps, and finally we end with $(b^{n+1}-w^{n+1})/(n+1)$. Adding all these contributions, we obtain $(b^{n+1}-a^{n+1})/(n+1)$ for the sum, the intermediate contributions canceling in pairs. From this behavior we may conclude that the average is exact, and, so far as accuracy is concerned, nothing is gained by subdividing the interval. Thus if we want the area under the curve x^n between $x=a$ and $x=b$, we simply multiply the right-hand member of Eq. (1) by $b-a$. Similarly, if we want the area under the curve x^{-n} between $x=a$ and $x=b$, we multiply the right-hand side of Eq. (2) by $b-a$.

At the outset these rules may be illustrated by application to cases with which the students are sufficiently familiar to know the answers. Cones and pyramids, as previously suggested, are such examples. It is also interesting in this connection to analyze the formula for the area in the annulus of a circle. The same principles may be used to obtain the volume of a thick spherical shell. In this latter case, the argument is essentially similar to that for the volume of a pyramid or cone, for it involves averaging a quadratic function of the variable. The area of a spherical surface of radius r is $4\pi r^2$. The volume of the spherical shell lying between $r=a$ and $r=b$ is the average of $4\pi r^2$ between these limits, multiplied by the thickness of the shell $b-a$. Thus we have for the required volume,

$$V=(4\pi/3)(b^2+ab+a^2)(b-a)=(4\pi/3)(b^3-a^3).$$

The application of the foregoing rules to a physical problem may be illustrated further by calculating the mass and moment of inertia, about its axis, of a thick cylindrical shell of axial length L, inner radius a and outer radius b. The mass of the shell lying between the cylindrical surfaces of radii a and b will be the average radius between a and b multiplied by $2\pi\rho L(b-a)$, where ρ represents the density of the material. The average radius is $(b+a)/2$ and the total mass M is $\pi\rho L(b-a)(b+a)$ or $\pi\rho L(b^2-a^2)$.

To determine the moment of inertia of the thick shell upon its own axis, we must multiply the average cube of the radius between a and b by $2\pi\rho L(b-a)$. The average cube of the radius between b and a may be represented by $(b^3+b^2a+ba^2+a^3)/4$. Thus the moment of inertia is given by $I=\pi\rho L(b^4-a^4)/2$. Insertion of the previously obtained value of M gives the familiar expression, $I=M(b^2+a^2)/2$,

for the moment of inertia of a cylindrical shell about its axis.

As suggested in this discussion, the average value may be introduced in several ways. If the exactness of the average is not previously established, a summation should be carried through. The advantage of using the algebraic expression for the exact average is then the simplicity of the resulting sums. If the exactness of the expression for the average has been justified to the satisfaction of the student, no summation is necessary and the solution of the problem is obtained directly.

The introduction of the average value appears to be justified by two considerations. First, the average value is a physically significant quantity which is not infrequently required in a problem; an algebraic expression for the average is the most direct approach to such problems. Secondly, the introduction of the average value puts beginners in a position to work problems in a way that is most natural to them. They are acquainted with the procedure used when constant quantities are involved. When varying quantities are introduced, they sense that the familiar methods might be used if only they could write down the averages. Giving a rule for finding the average is likely to provide the most direct answers to questions in the minds of students.

The algebraic expression of the average is convenient only when n is an integer. This is the case in most of the problems encountered in elementary physics. The method therefore provides a means of rationalizing in the minds of students many of the important results of physics without consuming an excessive amount of time on incidental routine.
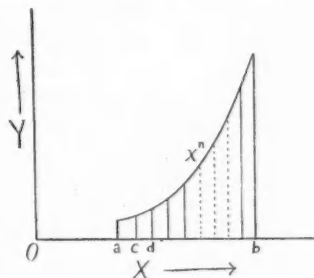

FIG. 1. Area under a curve.

Exact Averages for Use in the First-Year Physics Course

RALPH HOYT BACON

Department of Physics, New York University, Washington Square, New York City

EXTENSION OF COPELAND'S METHOD TO INCLUDE ALL EXPONENTS

AS Copeland points out,[1] his method for finding the average value of x^n in the definite interval from a to b fails when n is a fraction or -1. However, it is possible to devise an algebraic demonstration of the theorem for all values of n if we are willing to use a method that involves the process of passing to the limit, a process that Copeland's method avoids. To do this, we need the relations,[2]

$$\underset{m\to\infty}{\text{Lim}} \sum m/m^2 = \tfrac{1}{2}, \quad \underset{m\to\infty}{\text{Lim}} \sum m^3/m^4 = \tfrac{1}{4},$$

$$\underset{m\to\infty}{\text{Lim}} \sum m^2/m^3 = \tfrac{1}{3}, \quad \underset{m\to\infty}{\text{Lim}} \sum m^4/m^5 = \tfrac{1}{5},$$

and the well-known expansion

$$(1-z)^n = 1 - nz + \frac{n(n-1)}{2!}z^2 - \frac{n(n-1)(n-2)}{3!}z^3 + \cdots.$$

When n is -1, this reduces to

$$(1-z)^{-1} = 1 + z + z^2 + z^3 + \cdots.$$

We shall now determine the value of $\sum_a^b x^n \Delta x$. First, divide the range of x into m small intervals, where m is a very large number. To avoid questions concerning the convergence of the expansions used, we shall count these intervals from the upper limit of the range of x. Then $\Delta x = (b-a)/m$, and the value of x in the kth interval from b is $b - k(b-a)/m$. Thus, for $x^n \Delta x$, $\sum x^n \Delta x$ and $(x_{ab}{}^n)_{Av}$, we have the relations given in Table I. We can successfully substitute -1 for n in any step in the general expansion except Eqs. (6) or (7); these last two equations become indeterminate if this substitution is made in them.

Of course, in actual use, one would not, in general, proceed to Eq. (7), but would substitute the given quantities in Eq. (6) in order to solve the problem at hand. Among the problems soluble by this method are the calculation of the work done by an expanding gas, and of the

magnetic field due to a current in a straight wire (finite or infinitely long), by superposing the fields due to single moving charges.

EXACT AVERAGES OF SINE AND COSINE FUNCTIONS

An important exact average met with in the elementary physics course is the power in an alternating-current circuit. That the factor $\tfrac{1}{2}$ is exact may be easily demonstrated to students who have not had the calculus. To do this, we need the following formulas, all of which may be found in any elementary textbook on trigonometry:

$$\cos\phi_2 - \cos\phi_1 = -2\sin\tfrac{1}{2}(\phi_2+\phi_1)\sin\tfrac{1}{2}(\phi_2-\phi_1), \quad (1)$$

$$\sin\phi_2 - \sin\phi_1 = 2\cos\tfrac{1}{2}(\phi_2+\phi_1)\sin\tfrac{1}{2}(\phi_2-\phi_1), \quad (2)$$

$$\cos^2\phi = \tfrac{1}{2}(1+\cos 2\phi), \quad (3)$$

$$\sin^2\phi = \tfrac{1}{2}(1-\cos 2\phi). \quad (4)$$

Now, let the difference $\phi_2 - \phi_1$ be a very small angle $\Delta\phi$. Then the sine of the angle will be equal to the angle itself (easily shown and used several times in the physics course anyway), and Eqs. (1) and (2) become

$$\Delta(\cos\phi) = -\sin\phi\Delta\phi, \quad (5)$$

$$\Delta(\sin\phi) = \cos\phi\Delta\phi. \quad (6)$$

Eqs. (5) and (6) are of use in the "model dynamo" experiment.

Now let the long interval $\phi_b - \phi_a$ be divided into a large number of very small intervals. It is not necessary, in this case, that all these intervals be alike. Let us multiply the sine of the average value of ϕ in each interval by the width of that interval, $\Delta\phi$, and then add all these terms. Then, from Eq. (5), we obtain

$$\sum_a^b \sin\phi\Delta\phi = -\sum\Delta(\cos\phi) = \cos\phi_a - \cos\phi_b. \quad (7)$$

Similarly, from Eq. (6), we obtain

$$\sum_a^b \cos\phi\Delta\phi = \sum\Delta(\sin\phi) = \sin\phi_b - \sin\phi_a. \quad (8)$$

By drawing the graph of the function one can easily show that the sine or cosine of any argument averages out to zero over a sufficiently

---

[1] In the preceding article.

[2] Bacon, Am. Phys. Teacher **6**, 72 (1938).

TABLE I.

For $n \neq -1$	For $n = -1$
$$x^n \Delta x = \frac{b-a}{m}\left(b - \frac{k(b-a)}{m}\right)^n \quad (1)$$	$$x^{-1} \Delta x = \frac{b-a}{m}\left(b - \frac{k(b-a)}{m}\right)^{-1} \quad (1')$$
$$= \frac{b^n(b-a)}{m}\left(1 - \frac{k}{m}\frac{b-a}{b}\right)^n \quad (2)$$	$$= \frac{b-a}{bm}\left(1 - \frac{k}{m}\frac{b-a}{b}\right)^{-1} \quad (2')$$
$$= \frac{b^n(b-a)}{m}\left[1 - \frac{k}{m}n\frac{b-a}{b}\right.$$ $$\left. + \left(\frac{k}{m}\right)^2 \frac{n(n-1)}{2!}\left(\frac{b-a}{b}\right)^2 - \cdots \right] \quad (3)$$	$$= \frac{b-a}{bm}\left[1 + \frac{k}{m}\frac{b-a}{b}\right.$$ $$\left. + \left(\frac{k}{m}\right)^2\left(\frac{b-a}{b}\right)^2 + \cdots \right] \quad (3')$$
$$\sum x^n \Delta x = b^{n+1}\frac{b-a}{b}\left[1 - \frac{n}{2}\frac{b-a}{b}\right.$$ $$\left. + \frac{n(n-1)}{3!}\left(\frac{b-a}{b}\right)^2 - \cdots \right] \quad (4)$$	$$\sum x^{-1}\Delta x = \frac{b-a}{b}\left[1 + \frac{1}{2}\frac{b-a}{b} + \frac{1}{3}\left(\frac{b-a}{b}\right)^2 + \cdots \right] \quad (4')$$
$$= \frac{b^{n+1}}{n+1}\left[(n+1)\frac{b-a}{b} - \frac{(n+1)n}{2!}\left(\frac{b-a}{b}\right)^2\right.$$ $$\left. + \frac{(n+1)n(n-1)}{3!}\left(\frac{b-a}{b}\right)^3 - \cdots \right] \quad (5)$$	$$= \frac{b-a}{b} + \frac{1}{2}\left(\frac{b-a}{b}\right)^2 + \frac{1}{3}\left(\frac{b-a}{b}\right)^3 + \cdots \quad (5')$$
$$= \frac{b^{n+1}}{n+1}\left[1 - \left(1 - \frac{b-a}{b}\right)^{n+1}\right] = \frac{b^{n+1} - a^{n+1}}{n+1}. \quad (6)$$	$$= \log b/a = \log b - \log a. \quad (6')$$
$$(x_{ab}{}^n)_{Av} = \frac{\sum x^n \Delta x}{\sum \Delta x} = \frac{(b^{n+1} - a^{n+1})}{(n+1)(b-a)}. \quad (7)$$	$$(x_{ab}{}^{-1})_{Av} = \frac{\sum x^{-1}\Delta x}{\sum \Delta x} = \frac{(\log b - \log a)}{(b-a)}. \quad (7')$$
Q.E.D.	Q.E.D.

long range. Or, referring to Eqs. (7) and (8), we have

$$
\begin{aligned}
(\sin m\phi_{ab})_{Av} &= \sum \sin m\phi \Delta\phi / \sum \Delta\phi \\
&= \sum \sin m\phi \Delta(m\phi) / m \sum \Delta\phi \\
&= (\cos m\phi_a - \cos m\phi_b)/m(\phi_b - \phi_a).
\end{aligned}
$$

The denominator of the second member may become indefinitely large, whereas the numerator can never exceed 2. The second member, therefore, ultimately vanishes, leaving

$$(\sin m\phi)_{Av} = 0. \quad (9)$$

Similarly,

$$(\cos m\phi)_{Av} = 0. \quad (10)$$

To find the average value of the square of the sine or of the cosine, we use Eqs. (3) and (4), and proceed as before; thus

$$
\begin{aligned}
(\cos^2 \phi_{ab})_{Av} &= \sum \cos^2 \phi \Delta\phi / \sum \Delta\phi \\
&= \sum (1 + \cos 2\phi)\Delta\phi / 2\sum \Delta\phi \\
&= \tfrac{1}{2} + (\sin 2\phi_b - \sin 2\phi_a)/4(\phi_b - \phi_a).
\end{aligned}
$$

Following the same reasoning as above, we eventually obtain

$$(\cos^2 \phi)_{Av} = \tfrac{1}{2} \quad (11)$$

and

$$(\sin^2 \phi)_{Av} = \tfrac{1}{2}. \quad (12)$$

Now, to find the average value of the power in an alternating-current circuit, we have merely to find the average value of the square of the sine (or of the cosine) over a long period containing several cycles. Here, $\phi = 2\pi ft$, and

$$
\begin{aligned}
(\cos^2 \phi)_{Av} &= \sum \cos^2 \phi \Delta t / \sum \Delta t \\
&= \sum \cos^2 \phi \Delta\phi / \sum \Delta\phi = \tfrac{1}{2}.
\end{aligned}
$$

In case E and I are not in phase, let

$$
\begin{aligned}
E &= E_0 \sin 2\pi ft = E_0 \sin \phi \\
I &= I_0 \sin (2\pi ft - \alpha) = I_0 \sin (\phi - \alpha) \\
P &= E_0 I_0 \sin \phi \sin (\phi - \alpha) \\
&= E_0 I_0 (\sin^2 \phi \cos \alpha - \sin \phi \cos \phi \sin \alpha) \\
\bar{P} &= E_0 I_0 (\cos \alpha \, (\sin^2 \phi)_{Av} - \sin \alpha \tfrac{1}{2}(\sin 2\phi)_{Av}) \\
&= \tfrac{1}{2} E_0 I_0 \cos \alpha.
\end{aligned}
\quad (13)
$$

Physics and Society—A Symposium

THE papers by Walter S. Rogers, Herbert Feigl, F. M. Dawson and Robert E. Park which appear on the following pages were presented in a symposium on the important and timely topic, "Physics and Society," held at the State University of Iowa in connection with the fourth summer Colloquium for College Physicists, on June 15 and 16, 1939. The papers provide interesting interpretations, from several extra-physical points of view, of the significance of physics in our culture, and of the interrelations of the science with other important areas of knowledge.

This symposium was a most appropriate accompaniment to the unveiling exercises, in the physics library, of a mural entitled *Physics and Society*, which has been presented to the Iowa Department of Physics in memory of E. O. Dietrich, B.A. 1912, M.S. 1914, Ph.D. 1916, State University of Iowa. A photograph of the mural, with a description by the artist, Richard F. Gates, appears on this page. In the mural one may find a beautiful artistic expression of an idea voiced many years ago by the great physiologist and teacher, Emil Du Bois-Reymond—that "the history of science is the real history of mankind," the real record of human achievement.

It is to be hoped that the Iowa Colloquium for College Physicists will be continued from year to year, for there is a need for summer meetings conducted in an atmosphere that encourages serious discussion of the fundamentally inseparable problems of increasing physical knowledge and disseminating it.—D. R.

THE E. O. DIETERICH MEMORIAL, PHYSICS BUILDING, STATE UNIVERSITY OF IOWA

The E. O. Dieterich Memorial mural, *Physics and Society*, expresses in necessarily symbolic terms, man's conquest of the forces of nature, and particularly of electrical energy.

At the extreme left is represented primitive man, cowering in mad terror from the tremendous force which in his crudely anthropomorphic imagination must have seemed the anger of some hostile god. The forge-of-Vulcan group represents a more advanced theosophic expression of natural phenomena, while the mediaevel scholastic threatens with damnation any who doubt the thunderstorm as a direct indication of the Lord's wrath.

The cloaked man facing the approaching storm represents the beginnings, contemporary with the Renaissance, of direct observation and analysis, while the central group suggests modern laboratory method, supplemented by the work of the theoretical physicist to its upper right. The figure examining the sphere is intended to contrast the modern scientific approach, depending primarily upon examination of the material world, with that of the earlier scholastic, based on a stultifying traditionalism.

Two workers hold a modern city, representing in its complexity a thousand applications of man's new understanding of natural forces, while in the background a dam furnishes power for the city and irrigation for arid country-sides.

The climbing figures suggest man's realization that his present situation represents less a conquest completed than a struggle well begun; he proceeds with courage and determination toward a more complete understanding of the world in which he lives.—RICHARD F. GATES.

On the Significance of Science and Technology in Current World Affairs

WALTER S. ROGERS

Director of the Institute of Current World Affairs, New York, New York

ACCORDING to Nasr din Hodja, at a Friday service in the mosque the preacher asked those present whether they knew about a certain subject. They shook their heads. "That being the case," said the preacher, "you have taken no interest in it, so there is no point in my talking about it." And he dismissed them. On the following Friday he proffered another subject. This time thinking to outwit the preacher, all those present nodded their heads. "Well, if you already know about it, there is no point in my talking about it," said the preacher. Again he dismissed them. A week later, when asked as to their knowledge of a third subject, the members of the congregation were ready. Half of them shook their heads and half nodded. "Good," said the preacher, "let those who know, tell those who don't know." And again he dismissed them.

On several occasions I have wondered as to what might have been in the mind of that preacher. There are several possibilities. One of them is that he may have been in the predicament I find myself in now. I have in mind certain subjects which I wish to bring to your attention, but I do not know enough about them to talk with authority, and I have no desire to be merely glib.

My interest in the particular subjects I have now in mind arose out of my being one of the trustees of a sizable fund, the income from which is to be used to enable a few young men to gather and disseminate information, in this country and abroad, on current affairs, conditions, trends and personalities. Odd as it may seem, there are plenty of young men who—expenses paid—are willing to let themselves be persuaded to look into current problems, here and elsewhere, and, seemingly, there is no dearth of problems. So to spend the income of the fund appeared to the trustees to be easy enough; yet, odd as it may seem, sometimes even trustees can go looking for trouble. In this instance, one of the trustees gave expression to the feeling of the board by saying: "Let's take on a perpetual

headache, do some pioneering and see if we can spend the money in ways that will bring out fundamentals and lead to something constructive." The acceptance of that attitude led to a consideration of what are the forces at work in the world that are causing the present tensions, here and abroad, and apparently bringing about fundamental changes in institutions, relationships and ways of individual life, and a consideration of what are the agencies, personal and institutional, through which such forces are generated and made operative.

As the issues raised by such an approach are fairly simple, requiring for answer merely an apprehension of the nature of man and society, past, present and potential, I will not dawdle over them; like the preacher, I will leave their elaboration to the members of the congregation.

Instead, I will start with a mere suspicion that arose—a suspicion that perhaps some of the forces at work in the world stem out from modern science and technology. If this be more than a suspicion, and maybe it is, it seems pertinent to ask what are physics, chemistry, biology, engineering, scientific agriculture and medicine doing to the world? Or, if you prefer, what are physicists, chemists, biologists, scientific agriculturalists, doctors and their ilk doing, directly and indirectly, to you and me and to the institutions which we share, modify and create? And then, just for ample measure, what are we and our institutions doing to scientists, chemists, *et al.*?

The operation of forces generated by science and technology can be viewed from any number of levels. I will pick one for a moment's reference. A letter recently received from a young man who is attending the Indian Civil Service Probationers Course in London, contains the statement that India, a country with half the area of the United States, has about doubled its population in the past seventy years—growing from a mere two hundred million to nearly four hundred million, and now increasing in numbers at the rate of six million a year—and the further

statement that one school of thought believes that the solution of India's population problems is to be found in increasing and intensifying industrialization. The increase in population is due, in part at least, to the application of science and technology, and the hope is based squarely on their utilization.

India is but a particular case. The future of Great Britain, Germany, France, the United States, Soviet Russia, Japan and other countries will depend in part upon the extent to which progress is made in scientific and technologic development, and upon the extent to which the resulting products, methods and ideas are utilized. The relative progress of these countries in such utilization has a direct bearing on their respective positions in international relations. For example, behind the obvious political rivalry between Great Britain and Germany there has long been, and there is now, a rivalry in scientific and technologic development and utilization, having both peacetime and wartime aspects.

Unique questions, too, arise in regard to each country. Soviet Russia, for example, happens to be the first country to put to the forefront of governmental activity the promotion of science and technology; its success or failure in that endeavor may prove to be a much more decisive factor in determining the future of the country than, for instance, the liquidation of a few thousand of the enemies of the present regime. In regard to Japan curiosity exists, in many quarters, as to whether or not the Japanese will prove to be creative in science and technology. If not, they will always be at least one lap behind those who are. If they are, the result may well prove to be decidedly different.

As to our own country, we, or rather some of us, have demonstrated scientific capacity and inventiveness, and we, as a whole, have shown a surprising willingness to take unto ourselves the resulting new products and new modes of life. We may not be as good at science and technology as we like to think we are, but we are good. However, there are observers who are wondering whether we have not developed rigidities—economic, political, legal, and institutional—that will in the long run slow up scientific and technologic activities and prevent progress that might readily be possible under more flexible conditions. Scientists and technologists here may yet see their activities curtailed and their morale weakened by the failure of the community to utilize fully, for the general welfare, obvious potentialities. Similar issues arise in other countries.

Questions of educational policy are involved such as the provision of scientifically and technically trained men—for research, teaching, public and other services—not only in normally adequate numbers but in numbers sufficient to provide a surplus that can be drawn upon in emergencies such as a sudden demand for increased armaments. In the race that is now on, certain countries are seriously handicapped by the absence of such a reserve. Some countries are sending into other countries scientifically and technologically trained personnel for purposes of political penetration.

Such questions, and many similar ones that come to mind with only mild cerebration, tempt to facile—and wordy—elaboration; but I will pass up the opportunity, for now I am really concerned with how to get such subjects reported on so that they may be brought to official and general attention. Detailed research into such problems is another matter.

My concern came about in this way. My colleagues and I, having discovered that there are strange forces about, hit on the idea that it might be worthwhile to finance several young men, each trained basically in a science or a technology, to enable them to report on the world from the viewpoint of the impact of their respective fields, the men to be given plenty of time enroute to widen out their interests into the fields of economics, sociology and politics. Such men need not be technicians. As we liked our idea, we began to look for men—and have not found them.

There appear to be two main reasons why they do not exist. Both reasons are known to the members of this congregation, but, lacking the self-control of the preacher, I am going to state them anyway. Each of the scientific and technologic fields has been increasingly broken down into specialties and, seemingly, few, if any, men are being turned out with solid grasp of broad, basic principles and fundamentals. Progress doubtless has been made through special-

ization. But there are needs other than scientific research and technical application. I have mentioned a particular one. There is, furthermore, the desirability of people generally, living in a culture such as ours, having a broad knowledge of science and technology, and an understanding of their social significance.

These are superficial aspects of a general problem. Any one, not deaf, dumb and blind, who goes about this country can scarcely fail to be impressed by the steady advancement in every field of detailed knowledge, or fail to become aware of the difficulties being encountered in bringing that knowledge to bear on concrete situations and on the formulation of policy by individuals, social groups, institutions, corporations and government. For example, for the successful operation of a modern farm, I am told, there should be utilized an amazing complex of advancing scientific, technical, economic and social knowledge. It may well be that a great need of our time is the integration of knowledge at every conceivable level, which on the practical level means capacity to focus a great variety of detailed knowledge on a given situation. For some of such activities a range of basic knowledge may be requisite rather than specific technical proficiency. I am told, by people who probably do not know what they are talking about, that now the only way for a student to get a broad knowledge of any of the major fields is to study in a small institution too poor to offer highly specialized courses.

The other reason why we have not found men of the type sought, lies in the fact that only in recent years have present-day scientists and technologists come to have an interest in the social bearings of what they are doing and in the need for a friendly environment if they are to function fully and freely. Few courses are available on the history of science and technology and none, so far as I am aware, on their social significance. There seems to have prevailed an easy assumption that "the unseen hand of progress can be relied upon to make all things work together for good."

This lack of concern with social bearings and interrelationships and this easy assumption may well have been of no moment under erstwhile conditions, but things are no longer as they used

to be. New problems have arisen. Here are a few trivial ones: How, in the future, are science and technology to be organized, financed and generally furthered? What changes in attitudes and institutions will be needed to make available to the mass of the population of this and other countries new products and the ideas associated with them? What controls are necessary to prevent the use of science and technology for destructive purposes? What effects will result from the rapid industrialization of several of the South American countries? Such issues have personal, local, national and world-wide bearings.

As a noninitiate I find myself perplexed by the current urge that leads distinguished scientists and technologists to strike off papers on aspects of the interrelations of science and society. I have read many of the papers; in most cases the offering seems to me to be a mere verbal melange, revealing neither research nor reporting. Few of the authors appear to sense the urgency of the issues. Few exhibit any real knowledge of the developments in fact and theory that are taking place on the frontiers of the several social sciences and of philosophy, and of the far-reaching changes taking place in society. Only one or two of them face the possibility that for the full fruition of science and technology there may be required basic changes in our currently accepted mores and institutions. Incidentally, social scientists, too, shy away from such a prospect. Does prudence play a part? I wonder. Well, whether these papers mean much or little, the problems remain, and life, flowing along like "ole man river," is going to deal with them.

The fact that a few people believe that world affairs might well be reported on by men grounded in science and technology, and that their efforts might give impetus to the making of intensive studies of the problems involved in the interrelations of science and society, is not a particularly persuasive reason why scientists and technologists generally should bestir themselves over the training of such men or over the problems. But there is an appealing reason. Their jobs and the conditions under which they work are involved—and what is more, they and their families must live in the society that is in

the making. Science and living will be conditioned by that society.

A few scientists have become seriously concerned. As a result of their urging, a notable development has been to bring the study of the problems of the interrelations of science and society into the frame of activities of the British and American Associations for the Advancement of Science. But there seems to me to be a movement of even more promise. It is the apparent emergency in each of the scientific and technologic fields of what can roughly be called a social philosophy of the field. For example, the dean of one of the great engineering schools told me recently of what he was doing to further the development of a social philosophy of engineering that would include conceptions of the part engineering and engineers could, should and might play in shaping and furthering the economic, institutional and cultural development of our country. He added that he was a left-wing engineer, thus implying, no doubt, that he was disposed to go as far as facts and human needs led.

The struggle in the medical profession over the socialization of medicine indicates that traditional conceptions are being challenged and being replaced by newer ones more consonant with medical possibilities and human needs. There may well emerge a social philosophy of medicine that will serve to stimulate the practician, add to his culture and integrate medicine with other human efforts.

Why should there not be a social philosophy of physics? of chemistry? of biology? of agriculture? Such philosophies would give to men working in the respective fields a social consciousness and would provide factual and conceptual material for use by others in dealing with problems running from local to international. A minor, though perhaps very important, effect would be the contribution such philosophies might make to the building of morale. Recently, a medical student said to me: "In the hospital there is a badly malformed infant whose life will be saved and who will become a burden to himself, his family and society, and there is a diabetic—a bad case—who by grace of the latest developments will recover and go out perhaps to breed other diabetics. Does it make sense?" An agriculture student exclaimed to me, "What the hell is the use of working to raise additional crops when hungry people can't afford to purchase them?" A chemistry researcher expressed his despair because a new product he had developed was suppressed for commercial reasons. Even in my limited contacts I keep running across such cases. Unless students and practicians see a social significance in what they are doing the flaming torch will die down into a flicker.

Furthermore, the morale and social attitudes of scientifically and technologically trained people, who constitute an increasing percentage of every population, may turn out to be determinative factors in national and international evolution. The part now actually being played, as citizens and leaders, here and elsewhere, by people so trained is a subject that I prefer to leave to the members of the congregation for curbstone arbitrament.

Thus far I have not particularized in regard to physics. I see no reason why I should. No doubt it is often difficult to detect any direct social significance in a particular piece of physical research, but presumably the research is carried on with a view to adding to the body of knowledge of the subject, and surely physics is of far-reaching social significance.

For some years I have had a growing belief that training in fundamental mathematics and the basic principles of science makes for capacity to deal with social and political problems. However that may be, one of the ablest interpreters of international affairs that I know is an honors graduate of Cambridge in mathematical science. He has a rare ability to break down a complex problem, to deal with the several variables in turn, and then to relate them. I wish I knew several young men with similar training and with a desire to report on public affairs.

Maybe the preacher in the mosque was wiser than I am for he used fewer words, and took up less time. Perhaps he had the hope that I now have, which is that those in the congregation who know, in addition to telling those who dont, will also tell me.

The Significance of Physics in Man's Philosophy

HERBERT FEIGL

Department of Philosophy, State University of Iowa, Iowa City, Iowa

IN more than one way physics has profoundly influenced our philosophic outlook. In its picture of the universe physics achieved a sweep and depth that the philosopher could scarcely anticipate. In fact, he finds it exceedingly difficult to keep abreast with the rapid development of physical discoveries and the breath-taking constructions in physical theory. In a different respect physics as a branch of scientific knowledge has become a most important subject of analysis for philosophy. The logical, methodological and even social-ethical lessons to be learned from such a study of physical science are of the greatest significance.

Let me begin with a few highlights of the *cosmological* message of physics. The Copernican revolution still remains the most incisive step toward the modern world-picture. The abandoning of the egocentric, anthropocentric, geocentric standpoint; the liberation and objectification achieved through putting man and his planet in his, cosmically very insignificant, place in the universe; the realization that a great deal of the observed phenomena may be due to the frame of reference and the subjective condition of the observer—all this is most characteristic of modern scientific method. The colorful world of immediate experience is merely the point of departure. By bold construction and inference a world picture has been erected in which the relation of the observer to the observed is continually clarified and revised. Copernicus, like the ancients, still believed in a finite and bounded universe, however, with the sun in its center. Giordano Bruno, who died the martyr's death for his heretical doctrine, gave us the conception of a spatially infinite universe, thus rendering meaningless the question concerning a center. To Newton this infinite space, as well as infinite time, were absolute and independently existing realities. The wave theory of light seemed to give more substance to this absolute framework of space and so, for 200 years, the conception of an *ether* was regarded as something like the incarnation of absolute space. But today we know that no observable facts could confirm the assumption of a stationary ether, and so we have become reconciled to a thoroughgoing relativity of space and time. Einstein, on this basis, developed some of the most surprising results that science has ever seen: the relativity of length, duration and simultaneity; the relativity of mass, the equivalence of mass and energy; furthermore, the geometrical explanation of gravity and the unrestricted relativity of all motions. Finally—and this is the crowning effect —Einstein gave us the theory of a finite but unbounded universe; a universe that is self-contained, with a finite yet ever expanding (or perhaps, oscillating) volume.

An even greater wealth of strange facts were uncovered in the microcosmos. The ancient idea of the atomic structure of matter gained new support through the experimental and theoretical findings of physics and chemistry in the last century. Dalton's theory of atoms and its development in chemistry, the kinetic theory of heat, radiations from rarefied gases, radioactivity, and, in our century, the mathematical analysis of spectra, the exploration of the interior of the atoms initiated by Rutherford, the quantum theory of the atom of Bohr—all have opened up astounding vistas of the world of the very small. Good old solid matter with which we are so familiar is explained as the sensory appearance of an underlying reality which we cannot visualize and which, according to present theory, consists of a most intricate structure of whirling electric corpuscles dynamically interacting with one another through electromagnetic fields—the whole process described by probability functions.

The theory of relativity showed that the dogmas of absolute space and time, of the conservation of mass and of the uniqueness of Euclidean geometry could no longer be upheld. Similarly, modern atomic and quantum physics upset the belief in the immutability of the chemical elements and the principle of continuity, and has cast serious doubt upon the strict validity even of the most general pre-

supposition of all traditional science—the principle of causality. It seems as if prediction of events on the atomic scale were restricted to statistical averages and as if the individual occurrences could thus be determined only with a degree of probability. In connection with this development physicists have come to speak of the breakdown of the mechanistic world view. This seems to mean that the rigid determinism as well as certain other features of classical mechanics have to be relinquished. The world is no longer conceived along the lines of a machine or clockwork but rather in terms of the much more flexible concepts of fields, systems, waves and probability functions. On this basis some physicists have seen hopes of rescuing the universe from the ultimate fate of the "heat death" and also of reconciling the very unmechanical features of organic life and of human behavior with physical theory. (A few have gone so far as to think that modern quantum theory has provided the basis for free will!)

From this very sketchy enumeration of some of the most significant discoveries of modern physics it must seem as if the physicists have left the philosophers without a job. The old riddles of space and time, matter and force, cause and effect, etc., which at all times preoccupied the philosophers' attention, seem to have been taken out of their hands and put into those of the expert and the specialist. Einstein, in his Princeton lectures (1921), declared that he had to take down from the lofty Olympus of philosophic ideas the worn-out concepts of space and time and by radical repairs put them again into serviceable shape for use in physics.

The lesson that philosophy is slowly beginning to learn from all this is simple enough. It is not the business of the philosopher but that of the scientist to discover the truths about matters of fact—even if they concern the structure of the universe or man's place in it. Scientific progress has by far outrun philosophic speculation. The metaphysical systems and cosmologies of the classical philosopher look pale and insignificant when compared with the account of the world given by modern science. The philosopher can at best vaguely anticipate some of the grandiose scientific syntheses, but he is lucky indeed if his conjectures are at all in harmony with established fact.

Even historically speaking, philosophy has not been exclusively speculative. Ever since Socrates embarrassed the bright young men of Athens with such questions as: "What do you mean?" and "How do you know?" philosophy has, at least in part, pursued the task of *making our ideas clear* and on this basis examined the justification of our claims for knowledge. Thus, turning away from the precarious ambition to decipher the riddles of nature by *a priori* deduction or questionable inductive guesses, philosophy discovers its true nature in the *pursuit of meaning* and the *analysis of the methods of knowing*. Physics, being the most advanced, most systematic and most exact among the sciences of nature becomes a fascinating subject matter for the analytical or critical philosopher. Consider, for example, the dispute concerning the aim of science. Is it *description* (the answering of the question "how?") or is it *explanation* of nature (the answering of the question "why?"). Toughminded experimentalists maintain that science can only describe, that is, classify, relate and organize facts, but that it can never explain them. Galileo, to whom, more than to any other one searcher for truth, we owe the basic attitude of modern science, relegated the question "why" to the limbo of metaphysics and theology. We find the same spirit in Newton's "hypotheses *non fingo*" and in the works of the positivistic physicists of the nineteenth century. And yet a little philosophic clarification reveals that there is a good meaning for the term *explanation*—in common life, as well as in science —that implies no reference at all to ultimate reasons or purposes. This good meaning of *explanation* is perhaps best stated as the logical derivation of more specific (more descriptive) statements from more general (more constructive) assumptions. A pure description, strictly speaking, merely states a single fact (or groups of single facts) completely free from any interpretation. Such interpretation comes in only when we logically relate descriptions to laws and assumptions. If the specific case, for example, a phenomenon of light reflection, is deduced from the laws of geometrical optics, we say we *understand* it, or we have explained it. What we

mean here by *explaining* and *understanding* seems, therefore, nothing but the reduction of a special instance to a class of instances, whereby this class of instances is characterized by a generally valid law. But, interestingly enough, the laws of geometrical optics, together with the laws of diffraction, polarization, interference, etc., are in turn deducible from a set of assumptions concerning the nature of light—the wave theory. If we choose Maxwell's electromagnetic wave theory we recognize immediately its enormous explanatory power. A number of originally unrelated facts are synthesized into one harmonious system. Optics becomes a special case of electromagnetics. From this example, as well as from many others, we learn that there are *levels* of explanation. They can be conveniently arranged in the following way. On the ground level, as it were, we have pure description. Immediately above it we find the level of empirical or experimental laws. These laws in turn are explained by theories and it is often useful to distinguish two or even three levels among the theories. In thermodynamics, for example, we should have on the ground level any description of a thermal phenomenon, say, the cooling of a gas upon rapid expansion. On the level of empirical laws we should then place, among many others, the laws of Boyle and of Charles. All these laws find their systematic place in the theoretical framework of classical thermodynamics, with its basic first, second and third laws. The whole of classical thermodynamics in turn becomes deducible (after the theoretical discoveries of Clausius, Maxwell, Kelvin, Boltzmann and Gibbs) from statistical thermodynamics and its assumptions concerning the molecular theory of heat. And, finally, the modern quantum theory adds perhaps another level in giving us an account of the details of molecular interaction and of thermal radiation. Similarly in astronomy. Beginning with the immediate descriptions of the apparent locations and motions of the stars ("spherical astronomy") we find the first valuable empirical generalizations in Kepler's three laws of planetary motion. From here the question "why" drives us to the next level—where Newton's mechanics and his law of gravitation provide a new basis of deduction. Gravitational force and mass, and inertial mass, etc., were the uninterpreted ideas on this level. But Einstein, from a still more elevated standpoint, was able to find interpretations for them and to deduce Newton's laws (with certain corrections and additions) from the much more comprehensive theory of general relativity.

Naturally, the question arises: "Whither knowledge?" Do we go on forever piling explanations on explanations? Or does this process come to a natural end? Strictly speaking, we do not know and we cannot know. But the evidence of past scientific progress is very encouraging. There is a remarkable convergence of scientific theories. The two dominating theories of our time: the Einstein theory of relativity and the Heisenberg-Schrödinger-Dirac theory of quanta are almost sufficient to deduce all the more specific disciplines and their laws. To be sure, newly discovered facts or well-known, but recalcitrant, facts may require still higher levels of explanation or, at least, incisive modifications of the present top level. Science will probably never arrive at anything like theoretical completeness. Yet the surprising reductions of thermodynamics to molecular mechanics, of optics to electrodynamics, of chemistry to the theory of quanta teach most clearly the following lesson: The aim of explanatory science consists in the comprehension of a maximum of observable facts (the "data") by means of a minimum of basic principles and concepts. Scientific explanation is, however, never ultimate or infallible, but always relative, and, so to speak, "valid only until further notice."

The recent revisions of basic physical concepts are of great interest to the philosopher. The general lesson seems to be that what was rather dogmatically considered to be "*a priori* true" reveals itself as generalization from common experience in the realm of medium-sized things and events. As soon as we try to extrapolate to the large-scale phenomena of astronomy or the small-scale phenomena of atomic physics those preconceptions no longer fit and, therefore, have to be extensively modified. The guiding principle of these revisions has been clearly stated in P. W. Bridgman's operationism: Assertions are meaningful in physics only if they correspond to empirically testable situations. Physical concepts must be defined so as to be definitely connected

with the operations of observation, measurement and experiment. To adopt operationism means to exclude *meta*physics from physics. It should act as a safeguard also against such pseudo-philosophic excursions as some of the great physicists themselves have recently engaged in.

This attitude of a scientific empiricism seems also very promising in regard to *social-ethical* problems. The good scientist is characterized by his open-mindedness. Scientific discovery is, after all, a matter of trial and error, fumble and success. There is no philosophic principle about which we can afford to be dogmatic. Every law of nature, every scientific theory must be kept open for retesting and revision. This is true notwithstanding the fact that some parts of science are, of course, very well established and confirmed, and that we, therefore, do not care to revise all assumptions all the time. If this impartial and fact-minded procedure were adopted also for social organizing and planning, for the revision of individual, national and international attitudes, considerable progress would be facilitated in the direction of greater security, stability and happiness of mankind.

Social Contributions of Physics

ROBERT E. PARK
Department of Sociology, University of Chicago, Chicago, Illinois

IN an address before the National Academy of Sciences a few months ago Sir William Bragg, referring to a question which, as he noted, has become "a subject of the first concern of the whole world," made this statement:

> The effect of science upon social relations and social conditions has become very great, and the gains are obvious. Yet science does not appear to be in all cases beneficient. It has become a matter of anxious consideration whether or no the increase in the knowledge of nature must necessarily bring evil as well as good.

Sir William's observation seems to reflect a general change in the attitude and temper of men of science and of the modern world, with respect to the relation of science to human welfare. His words may perhaps be construed as indicating that the long period of the secularization of knowledge which has marked the progress of modern science is drawing to a close, or that science, under some occult influence of totalitarian politics, has suddenly awakened to a new sense of its responsibility to society and of its relation to other aspects of our culture—philosophy and religion, for example.

I mention this because the question of what physics, or physical science, has done *for* society is not, after all, a scientific question. It is, in the form I have stated it, a question for a philosopher rather than a sociologist. The question would have been easier to answer had it been: What has science done *to* society?

The contributions of physics to society and social life naturally are most obvious in the field of technology. The modern world has passed successfully through the age of iron, steel and electricity and now it is struggling to adapt itself to an age in which the social structure is largely determined by the automobile, the tractor and the airplane. These mechanical devices have sometimes been devastating in their effects, but at the same time they have been regenerating; in every case, they have drawn closer and tighter that steadily expanding web of mutual interdependence we call *society*.

The statement that a tool is, from the functional point of view, a mere extension of the hand no longer suggests, as it once did, either the manner or the extent to which man's natural powers have been not only multiplied but correlated by technologic devices. The fact is, that with the aid of machinery, man has become less and less immediately dependent upon the natural resources of his environment, and has become more and more dependent upon a complex social organization, through which these resources are collected, manufactured and distributed. The result is that the food quest in the modern world has ceased to be what it once was—a search for food—and has become a hunt for a job.

It is, however, the great cities that have most completely transformed the human habitat and

imposed upon human beings the discipline of a more completely mechanized world. Perhaps that is why social problems in great cities have, in recent years, assumed a technologic character. They have become problems in social engineering. Every function of life, as it becomes rationalized, tends to be carried on with the assistance of some sort of machine.

The modern city has long since ceased to be what the peasant village was, an agglomeration of individual habitations. Rather, it is like the civilization of which it is the center and focus, a vast physical and institutional structure in which men live, like bees in a hive, under conditions such that their activities are considerably more regulated, regimented and conditioned than is likely to appear to the spectator or be perceived by the inhabitant.

In order to be a little more specific in regard to the social contributions of physical science, let me say a word concerning my conception of society. Assuming that any association in which individuals carry on a common life is a society, the types of association which we expect to find in a society of human beings are (1) territorial, (2) economic and (3) cultural.

The territorial order.—Geography and the territorial organization of society get their importance from the facts that social relations are determined largely by physical distances and that social stability is insured when human beings are rooted in the soil. On the other hand, the most drastic changes in society are likely to be those that involve mobility and, particularly, mass migrations of peoples. This is so true that Frederick Teggart, who has given much attention to the subject, believes that most of the great advances in civilization are due, directly or indirectly, to the migration of peoples and the catastrophic changes which have accompanied them.

From this point of view it seems that every technical device, from the wheelbarrow to the airplane, which provided a new and more effective means of locomotion, has, or should have, marked an epoch in history. It is said that every civilization carries in itself the seeds of its own destruction. Such seeds are likely to be the technical devices that introduce a new social order and usher out the old.

The economic or competitive order.—Living creatures not only attract and repel other physical objects but they compete. Economic relations, wherever they exist, are the products of competition; but competition also eventually brings about some sort of cooperation, which in human beings takes the form of an exchange of goods and services. The economic order is the product of trade. The market and the area over which exchange takes place mark the centers and limits of economic society.

It is obvious that technologic devices have profoundly affected economic relations. In improving the means of transportation they have progressively extended the limits of the world market and of economic society. They have made mass production and mass distribution possible and are directly responsible for the existence of the capitalistic system as we know it. Insofar as they have brought the divers peoples of the earth into a world-wide web of economic relations, technologic devices have laid the basis for a world-wide political society and eventually of a moral and cultural order that will include all mankind. Finally, civilization, as distinguished from local and tribal cultures, is the product of commerce and the incidental division of labor which commerce permits.

E. Elliot Smith, the English anthropologist, in an account of certain of our primitive contemporaries, says:

> They have, of course, no agriculture and no domesticated animals, except the dog in some cases. They do not build permanent houses, and at most make rough shelters. They would seem formerly to have gone about naked, as many of them still do. They are ignorant of pottery-making, and of metal-working. They have no social classes, and usually no organisation in clans or other similar social groupings. In fact, their condition can truly be defined as being practically devoid of social institutions. Many of them still live in this original way in natural family groups, such, for example, as are found among the gorillas and other anthropoid apes.[1]

We have only to compare the material and intellectual wealth of modern civilized peoples with material and intellectual poverty of primitive peoples to gain some concrete notion of what the physical sciences have contributed to society as we know it and to human welfare as

[1] Smith, *Human History* (W. W. Norton, 1929), p. 183.

we understand it. An adequate interpretation of any society, demands something more than a knowledge of its external aspects. To know primitive people we must be able not merely to describe but to understand them. In short, we must know how to communicate with them, and, so far as possible, to see the world from their own point of view.

The cultural order.—The thing we call society is, of course, something more than a population aggregate having a territorial configuration; something more than "a geographical expression," or even an association for the exchange of goods and service. Society, in the sense in which that term is applied to human beings, is characterized by the fact that it imposes upon the free play of economic and egoistic forces the restraints of a political and a moral order. Yet, custom, conventions and law, by which society exercises control over the individual, turn out finally to be the products of communication; and communication is, as Bridgman shrewdly remarks,[2] "a device by which one endeavors as far as possible to anticipate the probable future actions of his fellows and so put himself in a position to make the necessary preparations."

But communication is something more than Bridgman's description indicates. It is a social-psychological process by which one individual is enabled to assume, in some sense and to some degree, the attitudes and the point of view of another; it is the process by which a rational and moral order among men is substituted for one that is merely physiological and instinctive. Communication "spins a web of custom and mutual expectation which binds together social entities as diverse as the family group, a labor organization, or the haggling participants in a village market." Communication and competition seem to be the two fundamental processes or forms of interaction by which a social order is initiated and maintained among the individuals whose life in common is the life of society.

Communication is, on the whole, an integrating and socializing process. It creates the loyalties and understandings that make concerted and

[2] Bridgman, *The Intelligent Individual and Society* (Macmillan, 1938).

collective action possible. It is by communication that the great fund of knowledge which we call science is accumulated and transmitted. Science, in fact, may be regarded as the kind of knowledge that can be communicated, and the kind of knowledge that grows and becomes more abstract and more precise by transmission.

I have gone into some detail in my description of the role and function of communication because it is so obviously fundamental to the social process, and because the extensions and improvements which the physical sciences have made to the means of communication are so vital to the existence of society and civilization. Having said so much, it is scarcely necessary to point out the importance of the printing press, telegraph, telephone, phonograph, radio and cinema in bringing about the revolutionary changes in political and cultural life that are visibly going on about us. All these instrumentalities, separately and together, have made possible the most extraordinary concentrations of political power that the world has ever known. The fact becomes all the more significant as we observe the extent to which it has been possible, through means devised by science, to mobilize these vast imponderable forces and put them into the hands of single individuals or a group of individuals; for example, dictators and their privy counselors.

I have attempted to suggest briefly, the extent to which physics and the applications of physical science have entered into the warp and woof of present-day social life. I have sought to make some estimate of the material resources which science and technology have made available for the uses of the modern world. The question which finally poses itself is whether science, in awakening the vast energies that are resident in the material world, has not brought into existence forces which science cannot hope to control. This, I take it, was implied in the question as originally formulated by Sir William Bragg in the address from which I have quoted. The problem that remains to be solved, therefore, is how to create moral forces that will balance the forces which physical science has unleashed. Stated in this form, however, the problem, insofar as it is likely to be solved by science at all, is for the social rather than the physical sciences.

The Significance of Physics in Man's Physical Well-Being

F. M. DAWSON

College of Engineering, State University of Iowa, Iowa City, Iowa

I BELIEVE it is true that all science and the application of science which is the essence of engineering have their roots in the basic sciences of physics, mathematics and chemistry. Of these the greatest is physics. It is trite to say that civilization as we know it could not have developed without the physical sciences. Perhaps the best proof is to consider what would happen if all the methods of communication, of transportation, of modern agriculture, of the household, of sanitation and medicine, and practically all that we consider commonplace were abolished. Could you imagine such a world? To begin with, the population would have to be drastically reduced and I suspect that those of us in academic positions would be the first to be eliminated.

All the sciences are interdependent. This statement is not altogether at variance with the facts as others see them. Let me quote from a recent discussion on molecular biology:[1]

> Biology has always had to wait on the development and application of the physical sciences before it could go very far along the road. Without physical optics there would have been no microscope and until the perfection of the microscopes the biologist was in the main limited to what his unaided senses told him. It was by means of the compound microscope that the development of the cell theory was made possible in the nineteenth century. Similarly it is by means of the new tools and technics developed in many instances by the physical sciences that the door to a biology of molecules has only recently been opened. The fundamental interrelation of all knowledge was never more dramatically illustrated. The fallacy of the argument for a moratorium on research in physics and chemistry was never so adequately exposed. Biology and medicine can go just so far, and then they must wait for new ideas, fresh technics and more powerful instruments. These technics or instruments may well be developed in a chemical or physical laboratory, in an astronomical observatory, or in the brain of a mathematician. Neither the physicist nor the astronomer nor the mathematician is necessarily thinking of biology, nor is he necessarily interested in it. But success in pushing out the boundaries of knowledge in any one discipline is bound to have its repercussions in all.

[1] *Annual Report of Rockefeller Foundation*, 1939.

It is not my purpose to enlarge on the modern development of the ultramicroscope, the ultraviolet microscope, the electron microscope, the ultracentrifuge or "tagged atoms." These may be left for others more learned in the sciences to discuss.

In this paper I wish to call to your attention once more that engineering is fundamentally the application of science and that we can go no farther nor faster than the science of physics will permit. As a profession we progress slowly because we lag behind the basic fundamentals. Engineers apply to practical affairs that which the physicists give. Other sciences, it is true, contribute much; it is not fair to say that engineering is based wholly on physics in the narrow sense. But each of these other sciences is itself dependent on what the physicist has accomplished in the not too distant past.

It should be realized that in this area as well as in many other fields of science the whole may be greater than the sum of its parts. This is particularly true in the case of a living organism or organization. The ideals and ideas behind the assemblage of the component parts form part of the whole. Separate the individual parts as you may, the thought processes which caused the parts to come together are undoubtedly an integral part of the whole. We do not question this when dealing with life itself; as one writer phrased it, "continued fragmentation will not of itself necessarily reveal the true inner meaning of life processes."

As an example of the development of thought processes as applying to everyday affairs, let us consider a very commonplace and simple device—an ordinary screw thread. The invention has been ascribed to Archimedes; but in all probability it was used in a crude form long before his era, for we are told that he used the principle in connection with a machine for raising water, which is a relatively complicated application of the principle of the screw. Today we never stop to consider the many trials and mistakes made during the centuries in endeavor-

ing to make screws cheaper and better. Until the nineteenth century, the manufacture of threads on an iron bolt or a wooden pin was a crude process with hand tools. Then came the lathe, which permitted the rotation of a bar of metal and the use of a cutting tool that advanced at a constant speed, thus cutting a helicoidal groove. Next a solid die was used, which, in effect, multiplied the number of cutting edges. It was not until 1900 that the self-opening die head was invented; this was speedily followed by the addition of chasers and cutters which are automatically released when they have completed their cut, then back off at high speed ready to start a new screw on a new piece of metal automatically fed by other mechanisms. Today the principle of the screw is old but modern engineering has devised ways and means of making them in extraordinary quantity, with a degree of precision and a quality of material unbelievably good. We trust our lives and the lives of others to screws, and never consider the thought processes that went behind the making of the machines for making other machines which in turn make the screws that again are parts of new devices for making some other product, and so on *ad infinitum*. Today we can neither travel, drink, nor eat without relying on screws.

I have dwelt on this single illustration because of a desire to discourage the notion that all physics and its applications are necessarily complicated. Man's physical well-being is directly traceable to such methods as man himself can evolve for doing quickly and cheaply that which otherwise would require a long time or could not be done at all. Please do not mistake my meaning. I am not discussing man as a social animal or as a spiritual being, but simply assert that the human race has benefited tremendously from the application of physical laws. That we do not also make men over, changing evil men into righteous and kindly folk, should not be considered as failure on our part. Rather, it should be a challenge to the social scientists and to religious workers to do better the work that lies in a field which they have preempted for themselves. For many years we have been led to believe that if mankind only had release from poverty, all would be improved and society could rapidly change for the better.

We admit the unevenness of distribution; that some have more than others, a few have too much, and others too little of this world's goods. Yet it is apparent that actual hunger and starvation do not exist among the millions living in the central plains of the United States. In all the Americas the situation is so much better in this respect than in the Eastern and Near Eastern countries. We have a right to ask—Are we as a people better off than our brothers who live under less favorable circumstances? The answer will be found probably in the next one or two generations, when the effects of our changing philosophies are more fully realized.

The significance of physics in freeing man from superstitions and fears has not received the attention that it deserves. A superstitious mind cannot be a sound mind, able to determine clearly fact from fancy, and so it is quite proper that this aspect be mentioned in connection with man's physical well-being. A well-balanced mind is essential and, to aid in its development, a certain amount of plain skepticism must be inculcated. Things are not always what they seem. Teachers of scientific subjects must show their students that the apparent is often an illusion but that the observed phenomenon is explainable in terms of fundamental laws provided the laws are thoroughly understood. Students must be shown that it takes insight, time and patience to master primary causes. There are many ways of illustrating these ideas by means of lecture-experiments. An objective of all teaching should be to open the minds of students, not only to the necessity for proper observation and careful study, but so as to lead them to question many of their native fears or inhibitions. This is one of the great functions of physics in man's well-being.

In preparing this brief paper my thoughts continually wandered over the vast range of common everyday activities that have been influenced or transformed by physicists. These manifold ramifications led me to renew acquaintance once more with an old book, McCurdy's *Great Masters*, in which the author states, in the preface,

> The earliest biography of Leonardo, that in the book of Antonio Billi, ends with the words, "His spirit was ever devising new things".

He was the most versatile genius of the age of the Renaissance, and the more genius approaches the Universal, the less it can be seen save in section. His study of science was in inception that he might thereby know the structure of what he represented and the laws of its formation. It was continued independently of the artistic purpose and ultimately superseded it. The work in art of his later years was either undertaken as illustrating some principle, or as a compromise with necessity, which parted him perforce from the study of natural phenomena and primary causes.

Perhaps we should revise the opinion that the fine arts come into being after science has produced the leisure and opportunity for artistic development, and consider that science is greater than the arts and comes later in the development of humanity. Perhaps they go hand in hand, one leading and then the other. Certainly both are essential to our well-being.

This is the age of metals and machines. For centuries the simple principles of the wheel, roller, gear and cam, inclined plane, screw, turning shaft, springs and all other relatively simple devices have been combined and recombined to make machines of great usefulness. Then came electricity, the internal combustion engine, the vacuum tube and radioactivity. The list is legion. Yet of them all, consider the man working in fundamental physics who first dreamed that such things could be. Long before any practical use could be made of the internal combustion engine—that devil of our modern age—someone played with it in the laboratory. Haltingly, crudely he worked, none too certain what it all was about; yet, as with Leonardo da Vinci, "His spirit was ever devising new things."

A Small Optics Shop

G. G. KRETSCHMAR
Physics Department, Walla Walla College, College Place, Washington

SOME years ago we constructed equipment for an experiment on the velocity of light. The operation of polishing the small, stainless steel mirrors by hand proved to be so tedious that a small polishing machine was designed and built for the purpose. This polisher later proved to be of value in polishing a number of Corning glass filters and other surfaces. From this beginning we have gradually added other simple equipment and have gained some experience in polishing optical surfaces, particularly plane surfaces. No claim for originality is made for any of the devices. This description of our equipment and methods is given with the hope that it may prove useful to others who wish to develop optical polishing equipment for experimental work or in connection with optics courses.

For producing flat surfaces, relatively simple equipment is needed. In the grinding, a lapping disk for rough work and a more accurate surface for the fine grinding are needed. The lapping disk, which may be about 30 cm in diameter, is made by machining an iron casting, and mounting it to rotate in a horizontal plane at about

100 rev/min. Fig. 1 shows the lapping disk surrounded with an ordinary dish pan for catching the grinding material and water. There is a hole in the center of the pan with a short riser soldered in place to accommodate the shaft of the disk. For the fine grinding a flat cast-iron plate about 20×30 cm is used. This is ribbed on the bottom and the surface worked to a true flat. Grinding powder of about 220 grit is used on the lapping disk and very fine grit or "finish-

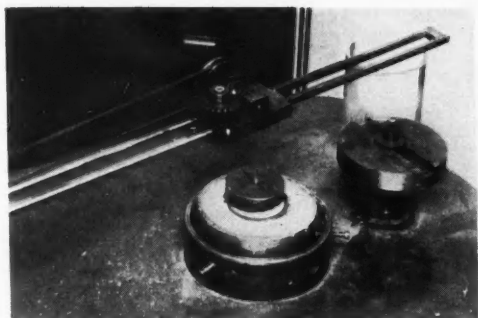FIG. 1. Lapping disk and surrounding pan.

FIG. 2. Head of the polishing machine.

ing emery"[1] is used on the plate for the finish grinding.

The polishing machine is of a conventional type, consisting of two rotating disks, arranged with movable driving pins so that both the length and the width of the stroke can be varied. Fig. 2 shows the head of the machine with the polishing arm raised from the polishing tool.

---

[1] This and the rouge for polishing can be obtained from a local manufacturing optician.

The material to be polished is held rigidly in place under the polishing tool and the latter is moved in a manner that may be described as an approximate one-to-one Lissajous pattern going slowly through its phase changes, due to a slight difference in the diameter of the two connecting drive pulleys. The tool is also slowly rotated by means of a worm gear on the polishing head.

Figure 3 shows the essential constructional features of the machine. The scale indicates the size of our equipment; but it is evident that similar equipment could be constructed in any necessary size. The machine is mounted on a small heavy table, and is driven by a small, variable speed gear motor. The motor is best mounted by means of springs so that relatively little vibration is transmitted to the table.

The polishing is done with beeswax tools that are backed with cast-iron disks about 8 mm thick. The disks which we use are of two sizes, approximately full size and $\frac{3}{4}$ size, as compared to the diameter of the work. The tools are made by melting a thick layer of wax on the iron disk, and then warming and pressing the wax until the surface is plane. Intersecting grooves are
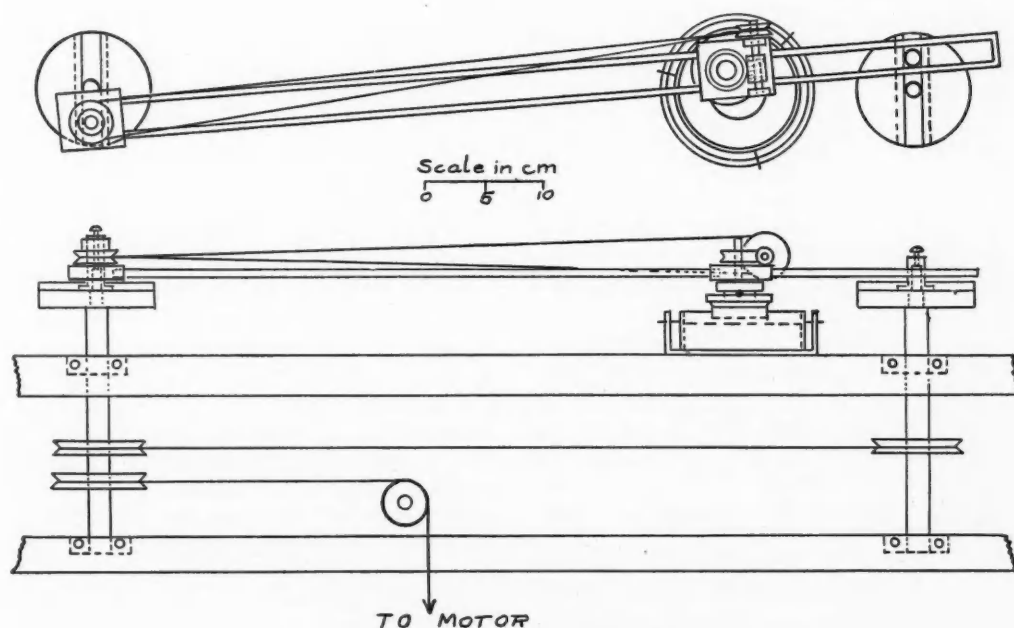

Scale in cm

TO MOTOR

FIG. 3. Constructional features of the polishing machine.

pressed deeply into the wax while it is warm, and the whole surface is finally trued by pressing it against plate glass. The tools are flexibly connected to the driving head by resting the latter on a single ball bearing, 8 mm in diameter. The turning motion is caused by two projecting pins which fit into a groove cut along one diameter of the disk.

The part to be polished is mounted by embedding it in plaster of Paris in the center of an iron ring. A convenient ring for the purpose is made from a 2-in. length of 4-in. standard iron pipe. To prevent the tool from turning down the edges of the work it is good practice to embed small supporting pieces symmetrically around the part to be polished. To set the glass so that the protecting pieces are flush with the main glass, the pieces of glass may be placed upside down on a flat surface in the desired pattern. The fine grinding surface is convenient for this purpose. It must be clean, and a thin film of oil will cause the surfaces to adhere when wrung into contact. The iron ring is then supported with spacers a little above the surface and the plaster of Paris, mixed rather thinly, is poured in, forming a layer about $\frac{1}{2}$-in. thick. As soon as the plaster has hardened enough to be firm the ring is slid off and the plaster is cut away to lower slightly the level from that of the surface to be polished.

After the plaster has hardened completely the glass in the ring must be fine-ground again to make sure that the surfaces are in the same plane. To know whether this final grinding is proceeding evenly, it is advisable to mark the surface lightly with pencil lines; when the grinding is proceeding uniformly over the surface, these marks should come off evenly.

Before proceeding with the polishing, the work is very carefully washed and every trace of grit removed. The work, still mounted in the iron ring, is then carefully centered and clamped under the head of the polishing machine. The rouge is kept in a stoppered bottle and mixed with water to about the consistency of cream. It is applied to the work with a brush and the work should not be allowed to run dry during the polishing operation. The polishing machine should be run as fast as practicable without undue vibration—say, 200 strokes/min. A short
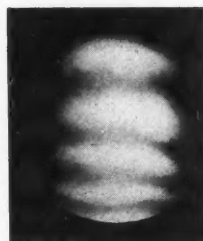

FIG. 4. Fringes produced between a pair of Fabry-Perot plates in contact.

stroke tends to polish the center low and a longer stroke tends to polish the edges low; so the procedure is to use the smaller tool with a short stroke to bring the center low, or to a polish first. Then, with the larger tool and a slightly longer stroke, the edge is brought down until testing shows the degree of flatness desired. The small tool will be used with a stroke that will overlap the work about 0.5 cm and the larger tool with a stroke that overlaps, say, 2 cm. As the work approaches a complete polish it is necessary to intersperse the polishing with frequent testing against a master flat under monochromatic light, using the well-known method of Newton ring fringes.[2] It will, of course, require experiment and experience to get flats easily, and the foregoing description is only suggestive of the general procedure.

For testing plane surfaces one needs a good optical flat. It is probably best to purchase it; however, good ones can be prepared by the well-known process of polishing three surfaces until they show the test for flatness in any combination of two.[3] Preliminary tests can be made against a selected piece of plate glass.

With the equipment described we have made a number of optical parts, including prisms and interferometer plates. Fig. 4 shows fringes obtained with a pair of plates that we produced.

In conclusion, I wish to express my appreciation to my assistant, Mr. Jamie Shumway, who worked out the procedure here suggested for figuring optical flats.

[2] See Strong, *Procedures in Experimental Physics* (Prentice-Hall), p. 64.
[3] See Michelson, *Studies in Optics* (Univ. of Chicago Press), p. 73.

The Use of an MKS System of Units in a First Course in Electricity

A. L. PATTERSON

Department of Physics, Bryn Mawr College, Bryn Mawr, Pennsylvania

THE adoption of mks units by the International Electrotechnical Commission and the interest aroused by the recent report of the Committee on Units of the American Association of Physics Teachers[1] makes it seem worth while to write an account of a first course in electricity and magnetism which has been given at Bryn Mawr College. It has been based on the definitive system of units developed by Campbell[2] from the original proposal of Georgi.

The meter, kilogram and second are used as the fundamental mechanical units, with the newton as the derived unit of force. In the present course the coulomb is chosen as the electrical unit with which to begin, since this choice results in a simplification of the dimensional formulas by avoiding fractional exponents. A straightforward qualitative account of electrostatics leads to the notions of charge, equality of charges, flow of charge, and the attraction and repulsion of charges. Coulomb's law is introduced experimentally with its suggestion that a unit of charge must be defined in order to determine the constant which appears therein. Demonstrations and laboratory experiments are considered for the purpose of defining a unit of charge that must be well defined, reproducible and permanent. Electrolysis is chosen to define the coulomb and the ampere, instead of using the so-called "absolute" definition which has been adopted by the I.E.C.

The volt is defined in terms of the joule and the coulomb, by means of the energy relations evidenced by the heating effect. The ohm is defined in terms of the ampere and the volt, through Ohm's law. The usual subsidiary concepts of resistivity, conductivity, etc., then appear, together with an elementary approach to Kirchhoff's laws. A qualitative discussion of the interactions of coils and wires carrying currents is given, and the notion of magnetism is intro-duced and given a tentative explanation on the hypothesis of atomic currents. The galvanometer is next discussed as an instrument that can be calibrated in terms of the units which have already been defined.

The notion of the capacitance of two conductors is developed from electrostatics, and the farad is defined in terms of the volt and coulomb. After a discussion of the properties of the parallel plate condenser of small separation, the law $C = kA/d$ is introduced experimentally.[3] The constant k (shown to be a property of the material between the plates) is defined as the dielectric constant with dimensions $[m^{-1}l^{-3}t^2q^2]$. The value k_0 for free space is then given as an experimental constant. The electric field strength is defined in volts per meter by means of the parallel plate condenser. The calculation connecting the constant k and the constant in Coulomb's law is described, the details being given in an appendix intended for those who are familiar with the calculus. The electrostatic voltmeter, electrometer, and the laws for condensers connected in parallel and series then follow. At this stage in the course a discussion of atomic structure is introduced, with a brief account of the determination of the properties of the elementary particles and the structure of the nucleus. The discussion of the atomic interpretation of the laws of electrolysis follows with an elementary account of primary and secondary cells and their properties. Contact potentials and thermoelectric phenomena are also touched on at this point.

The notion of equality of two magnetic fields is introduced by using a small coil carrying a current as a test object. The long solenoid is then investigated and the magnetic field strength is defined by the relation[3] $H = nI/l$.

The ballistic galvanometer is introduced and calibrated in terms of electromagnetic impulse. The permeability μ appears as an experimental constant of the medium in the equation that

[1] W. H. Michener *et al.*, Am. Phys. Teacher **6**, 144 (1938).
[2] *The Collected Papers of G. A. Campbell* (American Telephone and Telegraph Co., 1937), pp. 273–279; 466–477; 478–509; 510–521.

[3] It will be noted that "rationalized" units are used throughout.

connects the electromotive impulse induced in unit test coil with the change in magnetic field strength. The dimensions of μ are then $[mlq^{-2}]$. The magnetic flux density B is now introduced as equal to μH. The fact that a given rate of change in flux through a coil is always associated with the same electric field strength in the coil is then emphasized. Brief accounts of the simplest types of generators, of alternating currents, and of self- and mutual inductance follow. An account of the methods used in calculating magnetic fields due to currents is then given together with some of the results obtained. Ampere's law is introduced as a definition of H entirely equivalent to that already given.

The notions of electric displacement and displacement current follow from a discussion of a condenser with varying charge, and the connection between magnetic field strength and displacement current is obtained. This leads to a qualitative discussion of electromagnetic radiation, and of oscillatory circuits.

The relations between the mks units and the absolute units are given brief attention at the end of the course and a conversion table is provided to facilitate outside reading.

It will be seen that the course outlined gives a logical development of the theory of electricity and magnetism up to the point where a fairly sound descriptive treatment of electromagnetic waves is possible. Emphasis is laid throughout on the basic principles involved. Details of apparatus and of technical applications are taken up in the laboratory and, to some extent, by the suggestion of outside reading. We have found in our brief experience that, with the preparation described here, students are able to read more detailed accounts of technical applications than those usually given in first-year texts.

The course outlined forms a part of the first-year course in general physics, and occupies approximately three-fourths (10 weeks) of the second semester, with three hours of lecture and four hours of laboratory each week. It is now being given in essentially the same form for the second time. Our experience with the course has been very satisfactory. The use of one system of units, instead of the usual three, removes one of the worst sources of confusion from the mind of the student of electricity. When one system has been mastered, the problem of translating other systems into it offers little difficulty.

The laboratory program is greatly simplified by the present approach since the units in terms of which the familiar instruments are calibrated appear at the start. The formulation of realistic problems and their solution by the student is also facilitated.

In this course the material presented in the lectures has been supplemented by a set of mimeographed notes. As far as we are aware, the only text which presents a first course in electricity and magnetism based on an mks system of units is that of R. W. Pohl,[4] which is much too detailed for use in the present course. We wish, however, to express our indebtedness to this work, particularly in planning the later parts of the course. The intermediate texts of Vigoureux and Webb[5] and of Harnwell[6] will also prove valuable to anyone wishing to develop a course along similar lines.

In a more advanced course in electricity and magnetism given here last year, Professor W. C. Michels also made use of an mks system. He reports that it proved to be equally satisfactory. I wish at this time to express my thanks to him for many valuable discussions in connection with the development of this course, and to Miss Mary Carll, Miss Pauline Rolf and Miss Mary Donohue for their assistance in the laboratory and in the preparation of the notes. I also wish to express my great indebtedness to Dr. G. A. Campbell and Dr. Marion Gray for their advice and criticism.

[4] R. W. Pohl, *Die Elektrizitätslehre* (Springer, 1935, ed. 4).
[5] P. Vigoureux and C. E. Webb, *Principles of Electric and Magnetic Measurements* (Prentice-Hall, 1937).
[6] G. P. Harnwell, *Principles of Electricity* (McGraw-Hill, 1938).

The Forthcoming Columbus Meeting of the A. A. P. T.

Experiments with Acoustic Plane and Space Gratings

JAMES W. McGRATH*
State University of Iowa, Iowa City, Iowa

THIS paper is additive to the recent emphasis on an acoustic approach to the study of waves. Interference and diffraction can be demonstrated in a more direct perceptual way by means of acoustic, rather than optical, experiments. Such things as path differences may be seen and measured in manual fashion by the student. It seemed, therefore, advisable to design an acoustic spectrometer for use in the classroom and laboratory. It is described in its present imperfect form because of its suggestiveness with regard to new experiments.

The spectrometer can be operated in most fairly quiet rooms. The relative positions of the parts of the apparatus within a room should be judiciously chosen so as to lessen the effect of sound reflection, but sound absorbent drapes are not essential. These qualities make the spectrometer useful for demonstration and experiment.

The spectrometer base is 35 cm high and is composed of a vertical shaft mounted in a heavy tripod (Fig. 1). On this shaft there are three collars, separated by roller bearings, to which are attached the arms to the receiver and sound box, and also the turntable. The periphery of the turntable is graduated in degrees, and a pointer attached to each of the arms indicates the angle the arm makes with the grating plane. The arm extending to the sound box is so de-

FIG. 1. The acoustic spectrometer.

signed that the axis of the box intersects the axis of the turntable when the connection is made. A wooden base fastened to the turntable contains a slot which holds the front grating plane. Two ways are provided for the rear grating plane, and thus the interplanar distance can be varied continuously while the two planes are held parallel. The axis of the turntable passes through the center of the front plane; and while it need not, this seemed to be the optimum arrangement. The arms are about 125 cm in length; the one supporting the sound receiver is counterbalanced.

The gratings are made of steel. Each contains 7 flat elements, $31 \times 2.5 \times 0.3$ cm, and the distance between the elements is 10.2 cm. Round elements were tried but with less successful results. A further advantage of flat elements is that the phase changes across their faces may be more easily computed.

Auxiliary equipment includes the sound source (a Galton whistle is suitable) in a sound box. The box is $150 \times 36 \times 30$ cm in size and is similar to those described by Schilling.[1, 2] The receiver, which is very satisfactory, consists of a crystal microphone at the focus of a parabolic mirror. This reflector is necessary if one wishes to operate the spectrometer in a room where few acoustic precautions are taken. The response of the microphone was rendered visible by means of an oscillograph and amplifier.

To minimize the relative effect of sound reflection and general room noise, it is necessary to place the microphone and sound box close to the spectrometer. The distances used were about 70 cm.

Several factors limit the perfection of the response of this spectrometer; namely, (1) the grating elements and planes are few in number, (2) in the space grating, at some angles, the elements of the first plane hide those of the second, (3) the incident beam may not be sufficiently plane, and (4) the microphone also
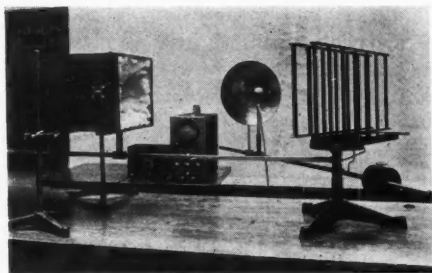
---

* Now of the Department of Physics, Michigan State College, East Lansing, Michigan.

[1] Schilling and Whitson, Am. Phys. Teacher **4**, 27 (1936).
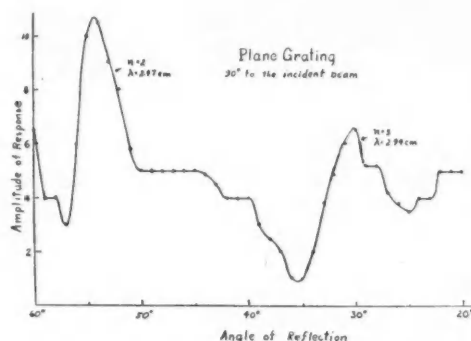[2] Schilling, Am. Phys. Teacher **6**, 156 (1938).

FIG. 2. Spectra from the plane grating.

responds somewhat to the rays that are not coming to the reflector in a parallel beam. Therefore, one must expect some interference effects other than those taken into account in the theory of an ideal grating. These extraneous effects do not appear strongly in the plane grating but are more pronounced with the space grating since the limitations there are more serious.

The following experiments have been carried out:

Plane grating at 90° to the incident beam.—This gives a satisfactory spectrum. A typical one is shown in Fig. 2. The first order does not appear, for the sound box interfered with the placing of the microphone. Since the angle read was that between the grating plane and the reflected beam, the law illustrated is $n\lambda = d \cos \theta$, where d is the distance between element centers. The student may actually measure the path differences and thereby see the grating's principle.

Space grating.—Because unwanted effects enter (mainly because only two planes are used) and because it is tedious to cover all angles, keeping the angles of incidence and reflection equal, only

TABLE I. *Average experimental values.*

METHOD	λ(CM)
Plane grating	2.95
Space grating, spectrum	3.16
Space grating, changing d	2.76

the regions where peak responses were to be expected were investigated. The peak contours are shown in Fig. 3. The formula demonstrated in this instance is $n\lambda = 2d \sin \theta$, where d is the plane separation. It is especially instructive to adjust the planes of the gratings and the microphone position so that the angles of incidence and reflection are 45°, and then observe the effect of varying the interplanar distance. Pronounced maximums and minimums of response of the microphone appear. In this way the student sees directly that it is the interference between planes that is utilized in the space grating and he may, by this method, measure the wave-length of the sound.
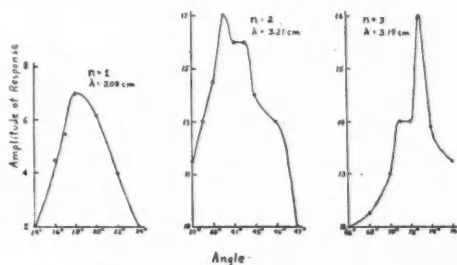
FIG. 3. Peaks of response from the space grating; $d = 5$ cm.

Although the average values from these experiments given in Table I show fairly good agreement, the various values obtained by each method agreed among themselves much more closely. No reason has suggested itself for the small differences in the results obtained by the three different methods, and hence an investigation of this point should be carried out.

Doubtless many improvements can be made on this spectrometer to give spectra more like those expected from an ideal grating. Any change that more nearly satisfies the best theoretical conditions (such as adding a third grating plane) should improve the performance. It is possible that smaller grating elements (these were about 1 wave-length wide) would produce more regular spectra. Many experiments other than those described here may be performed.

The Use of Exact Corrections for Damping in a Mutual Inductance Circuit

P. L. Bayley and Kenneth B. Shiffert

Department of Physics, Lehigh University, Bethlehem, Pennsylvania

IN the standard experiment on mutual induc-
tion, one attempts to show that, on reversal
of a given current in the primary circuit, the
quantity of electricity induced in the secondary
circuit is inversely proportional to the resistance
of the secondary circuit. A plot of the initial
throws of the ballistic galvanometer as a func-
tion of the reciprocals of the resistance of the
secondary circuit gives a curve such as A, Fig. 1.
As the resistance is lowered, the electromagnetic
damping is increased, which causes the deflec-
tions to be increasingly less than they should be.
If one determines the logarithmic decrement λ
for each value of the resistance and multiplies
each initial deflection by the appropriate factor
$1+(\lambda/2)$, he obtains a curve such as B, Fig. 1.
The experiment is limited to the range in which
$1+(\lambda/2)$ approximates the exact correction
factor. It is the purpose of this paper to show
how to extend the range of the experiment and
obtain a straight line such as C, Fig. 1.

Since the exact expression for the correction
factor is found in very few texts in English, it
seems worth while to outline the derivation here.
The solution of the general case, considering
that the emf induced in the secondary circuit
extends over an appreciable portion of the first
throw of the galvanometer, is given by Laws.[1]
With air-cored mutual inductances, one may
consider that the emf induced in the secondary
decreases to zero before the galvanometer has
moved appreciably. Under these conditions the
differential equation of motion of the gal-
vanometer coil is[2]

$$(d^2\theta/dt^2)+2b(d\theta/dt)+k^2\theta=0, \quad \text{(S. 337)}$$

where $2b$ involves the retarding torque due to the
currents induced in the moving coil and to air
friction (both proportional to the velocity of the
coil), and k^2 involves the restoring torque of the
twisted suspension.

The general solution of this equation is

$$\theta=A\epsilon^{[-b+(b^2-k^2)^{\frac{1}{2}}]t}+B\epsilon^{[-b-(b^2-k^2)^{\frac{1}{2}}]t}. \quad \text{(S. 337)}$$

If time is counted from the beginning of the
motion of the coil, then $t=0$ when $\theta=0$. Also, if
the coil gains an initial angular velocity ω_0, im-
parted by the surge in the secondary before
the coil has moved appreciably, then $t=0$
when $d\theta/dt=\omega_0$. From these two sets of initial
conditions we obtain $A=\omega_0/2(b^2-k^2)^{\frac{1}{2}}$ and
$B=-\omega_0/2(b^2-k^2)^{\frac{1}{2}}$. When $k>b$, which holds
up to critical damping, the solution of the equa-
tion of motion becomes

$$\theta=(\omega_0/(k^2-b^2)^{\frac{1}{2}})\epsilon^{-bt}\sin(k^2-b^2)^{\frac{1}{2}}t. \quad \text{(S. 337)}$$

The relation between the quantity of elec-
tricity Q passing through the galvanometer and
ω_0 is

$$\omega_0 J=HAQ, \quad \text{(S. 256)}$$

where J is the moment of inertia of the galvanom-
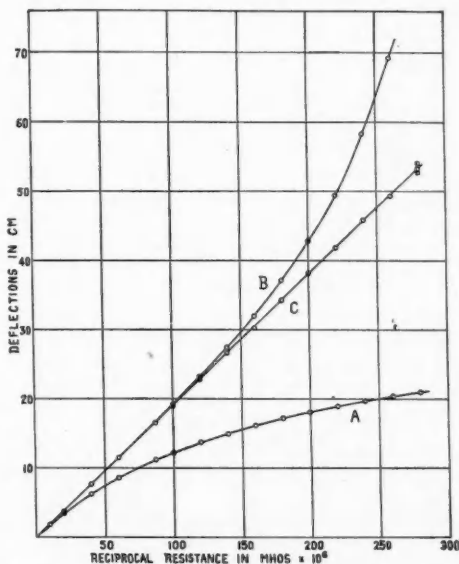
Fig. 1. Curve A: galvanometer deflections as a function
of the reciprocals of the resistance of the secondary circuit.
Curve B: the ordinates of Curve A multiplied by $1+(\lambda/2)$.
Curve C: the ordinates of Curve A multiplied by the exact
correction factor F.

[1] Laws, *Electrical Measurements* (McGraw-Hill, ed. 2),
p. 95.

[2] Equations here indicated by "S" will be found in many
intermediate texts on electricity and magnetism. The num-
bers following "S" refer specifically to pages in Starling's
Electricity and Magnetism (ed. 6).

eter coil, A is its total effective area and H is the strength of the magnetic field in which the coil rotates. Hence

$$\theta = [HAQ/J(k^2-b^2)^{\frac{1}{2}}]\epsilon^{-bt}\sin (k^2-b^2)^{\frac{1}{2}}t. \quad (1)$$

This function represents a sinusoidal motion whose amplitude decreases exponentially with time. Its zero values occur at regular intervals when $(k^2-b^2)^{\frac{1}{2}}t$ is any integral multiple of π. Hence,

$$(k^2-b^2)^{\frac{1}{2}} = 2\pi/T_\lambda, \quad (2)$$

where T_λ is the period of the damped motion. The condition for maximum values of θ, found by equating to zero the derivative of Eq. (1), is

$$\tan (k^2-b^2)^{\frac{1}{2}}t = (k^2-b^2)^{\frac{1}{2}}/b = 2\pi/bT_\lambda. \quad (3)$$

Noting that the arguments of the sine and tangent functions in Eqs. (1) and (3) are the same, we see that the period may be measured either from the time between the maxima or from the time between the zero values. Nevertheless, due to the factor ϵ^{-bt}, the maxima are displaced toward the origin from the midpoints between the zero crossings. Thus the first deflection θ_1 occurs at a time $t=t_1$, which is less than $T_\lambda/4$; that is, the phase angle $(k^2-b^2)^{\frac{1}{2}}t_1$ is less than 90°. When the decrement d (defined below) is 2, the first deflection occurs at about 78°, while when d is 3 it occurs at about 70°. Most approximate theories of damped harmonic motion assume that t_1 is a full quarter-period.

Let d be the ratio of any maximum deflection θ_n, occurring at time t_n, to the next maximum θ_{n+1}, occurring at time $t_n+(T_\lambda/2)$. From Eq. (1), $d = -\theta_n/\theta_{n+1} = \epsilon^{bT_\lambda/2}$. Hence

$$\lambda = \log_\epsilon d = bT_\lambda/2. \quad (4)$$

Equation (1) will now be modified so that the relation between Q and the first maximum θ_1 will involve only the galvanometer constants and λ. From Eqs. (3) and (4) we obtain

$$\tan (k^2-b^2)^{\frac{1}{2}}t = \pi/\lambda,$$
$$\sin (k^2-b^2)^{\frac{1}{2}}t = \pi/(\pi^2+\lambda^2)^{\frac{1}{2}}. \quad (5)$$

For the time t_1 we obtain from Eqs. (5) and (2),

$$t_1 = (T_\lambda/2\pi) \tan^{-1} (\pi/\lambda),$$
$$bt_1 = (\lambda/\pi) \tan^{-1} (\pi/\lambda), \quad (6)$$

where the smallest value of the arc tangent is to be used.

When $b=0$ there is no damping. Then, from Eq. (2), we get $2\pi/T_0 = k$, where T_0 is the period without damping. Combining this expression with Eqs. (3) and (4), we have

$$T_\lambda\pi = T_0(\pi^2+\lambda^2)^{\frac{1}{2}}. \quad (7)$$

Now when Eqs. (2), (5) and (6), and finally Eq. (7) are substituted in Eq. (1), we obtain the desired relation between the quantity of electricity passing through the galvanometer and the first deflection θ_1; namely,

$$Q = (2\pi J/HAT_0)[\epsilon^{(\lambda/\pi)\tan^{-1}(\pi/\lambda)}]\theta_1. \quad (8)$$

The expression in the brackets will be called the *exact correction factor F.*

In Laws' derivation[3] it appears that an additional expression $\pi/(\pi^2+\lambda^2)^{\frac{1}{2}}$ is included in the correction factor. However, one notices that the T used there is really T_λ, which involves λ, and which therefore must be included in the correction factor. When combined with Eq. (7), Laws' equation reduces to Eq. (8).

A brief table of values of the exact correction factor is given by F. Kohlrausch.[4] An expanded and modified version of it is given in Table I. As in Kohlrausch's table, values of F are given for equal steps of $\log_{10} d$ since this quantity is most easily obtained from a set of data. In preparing the table, every tenth value of F was computed with the help of 8-place logarithm tables. The intermediate values were obtained by using Newton's interpolation formula.

When it is assumed that the first maximum deflection occurs at the quarter-period, the correction factor should be exactly $d^{\frac{1}{4}} = \epsilon^{\lambda/2}$, of which $1+(\lambda/2)$ is an approximate value. A glance at the brief Table II shows that $d^{\frac{1}{4}}$ is actually a very poor correction factor while $1+(\lambda/2)$ is surprisingly good. The values of $1+(\lambda/2)$ at first are slightly less than F, reach a maximum difference at about $d=1.7$, become equal to F at about $d=2.24$ and finally become increasingly larger than F. Table III shows the percentage error caused by using $1+(\lambda/2)$ instead of F.

It is interesting to note that as the condition of critical damping is approached the exact correction factor approaches asymptotically the

[3] Reference 1, p. 102.
[4] Kohlrausch, *Praktische Physik* (B. G. Teubner, ed. 17), Appendix, Table 16.

TABLE I. *Corresponding values of* $F[= \epsilon^{(\lambda/\pi)\tan^{-1}(\pi/\lambda)}]$ *and* $\log_{10} d[= \lambda/\log_e 10]$.

LOG₁₀ d	F	LOG₁₀ d	F	LOG₁₀ d	F	LOG₁₀ d	F	LOG₁₀ d	F
0.000	1.00000	0.200	1.23235	0.400	1.45777	0.70	1.75540	1.50	2.25118
005	00576	05	23814	05	46319	71	76425	55	27048
10	01153	10	24392	10	46860	72	77303	60	28878
15	01730	15	24969	15	47399	73	78173	65	30612
20	02307	20	25547	20	47937	74	79036	70	32257
0.025	1.02886	0.225	1.26123	0.425	1.48473	0.75	1.79891	1.75	2.33817
30	03464	30	26699	30	49009	76	80739	80	35297
35	04043	35	27274	35	49542	77	81580	85	36702
40	04623	40	27849	40	50075	78	82413	90	38036
45	05203	45	28423	45	50606	79	83239	95	39302
0.050	1.05783	0.250	1.28997	0.450	1.51135	0.80	1.84058	2.00	2.40506
55	06364	55	29569	55	51663	81	84869	10	42739
60	06945	60	30141	60	52190	82	85673	20	44760
65	07526	65	30713	65	52715	83	86470	30	46593
70	08108	70	31283	70	53239	84	87259	40	48258
0.075	1.08690	0.275	1.31853	0.475	1.53761	0.85	1.88042	2.50	2.49775
80	09272	80	32422	80	54281	86	88816	60	51158
85	09854	85	32990	85	54801	87	89584	70	52423
90	10437	90	33557	90	55318	88	90344	80	53581
95	11019	95	34123	95	55834	89	91098	90	54643
0.100	1.11602	0.300	1.34689	0.500	1.56349	0.90	1.91844	3.00	2.55620
05	12185	05	35253	10	57373	91	92583	10	56519
10	12767	10	35817	20	58390	92	93314	20	57348
15	13350	15	36380	30	59402	93	94039	30	58115
20	13933	20	36941	40	60406	94	94757	40	58824
0.125	1.14516	0.325	1.37502	0.550	1.61404	0.95	1.95467	3.50	2.59482
30	15098	30	38061	60	62396	96	96171	60	60092
35	15681	35	38620	70	63380	97	96868	70	60660
40	16263	40	39177	80	64358	98	97558	80	61189
45	16846	45	39734	90	65329	99	98241	90	61682
0.150	1.17428	0.350	1.40289	0.600	1.66293	1.00	1.98917	4.00	2.62143
55	18010	55	40843	10	67250	05	2.02198	5.00	65444
60	18592	60	41396	20	68200	10	05315	10.00	70165
65	19173	65	41948	30	69142	15	08274	∞	2.71828
70	19755	70	42499	40	70078	20	11082		
0.175	1.20336	0.375	1.43048	0.650	1.71006	1.25	2.13745		
80	20916	80	43597	60	71928	30	16270		
85	21496	85	44144	70	72842	35	18663		
90	22076	90	44689	80	73748	40	20931		
95	22656	95	45234	90	74648	45	23081		
0.200	1.23235	0.400	1.45777	0.700	1.75540	1.50	2.25118		

value of ϵ while both $d^{\frac{1}{2}}$ and $1+(\lambda/2)$ approach infinity.

EXPERIMENT

A 2057-ohm Leeds & Northrup old-type HB galvanometer, a resistance box and one coil of a mutual inductance made up the secondary circuit. The galvanometer had a critical damping resistance of about 3570 ohms. A mutual inductance was selected that would give about 20 cm deflection for a reversal of 0.700 amp in the primary when the secondary resistance was low enough to cause critical damping. Then the resistance was raised until only 2-cm deflection

occurred. Values of the total secondary resistance were computed which would give equally spaced reciprocals so that the points plotted on Curve B, Fig. 1, would be uniformly distributed. Then, for

TABLE II.

LOG₁₀ d	d	$d^{\frac{1}{2}}$	$1+(\lambda/2)$	F
0.0	1.000	1.0000	1.0000	1.0000
.1	1.259	1.1220	1.1151	1.1160
.2	1.585	1.2589	1.2303	1.2324
.3	1.995	1.4125	1.3454	1.3469
.4	2.512	1.5812	1.4605	1.4578
.5	3.162	1.7783	1.5757	1.5635
.6	3.981	1.9953	1.6908	1.6630

TABLE III.

λ	Percent error by use of $1+(\lambda/2)$	d
0	0	1
0.27	−0.10	1.3
0.50	−0.15	1.7
0.70	−0.10	2.0
0.81	−0.00	2.2
0.87	+0.1	2.4
0.94	+0.2	2.6
1.05	+0.5	2.9
1.22	+1.0	3.2
1.46	+2.0	4.2
1.94	+5.0	7.0

reversal of 0.700 amp in the primary circuit, deflections were observed for each of the computed resistances. All deflections were taken alternately right and left to reduce errors due to zero shift. By the use of a previous calibration test, deflections were corrected for nonlinearity. The galvanometer resistance was measured by the Kelvin method, which avoids heating of the copper wire in the galvanometer coil.

When the secondary resistance was large, at least the first four successive deflections were recorded. The resistance was lowered, step by step, until the second deflection was a fraction of a millimeter and the third deflection was undetectable. Thus critical damping was closely approached. Although, under these limiting conditions, the error in the computed value of d was extremely large, little error is caused in F since, as has been pointed out, its values change very slowly as d becomes large. Thus, for the last point on Curve C, the first deflection was 20.96 cm and the second was 0.03 cm. The value of d was 699 and F was 2.540. Now, if it is assumed that the second deflection may have been in error by 0.01 cm and that the zero shift was a like amount, then the second deflection may have been 0.01 or 0.05 cm. For these data d would have been 2096 and 419, respectively, giving F values of 2.583 and 2.514, respectively. These corrections are indicated by squares on Curve C. They indicate the largest errors possible. For no other point of the curve could the uncertainty be nearly so large, because the second deflection would be larger and the percentage error in d thereby reduced greatly. Thus, for the second to last point on the curve the same assumed error gives F values of 2.434 and 2.411.

It has thus been shown that the linear relation between Q and the reciprocal of the secondary resistance can be easily demonstrated nearly up to the critical damping limit provided the exact correction factors are applied.

A Three-Dimensional Diagram of Gyroscopic Precession

Francis W. Sears

George Eastman Research Laboratory of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts

A S was recently pointed out by Loeb,[1] the explanation of gyroscopic precession in terms of the angular momentum vector and its changes is extremely unsatisfying to most students, whose usual comment is: "But I still don't see why it doesn't fall." The apparatus herein described represents an attempt to reduce the precessing gyroscope to a system of point masses, and to aid in visualizing the velocities and accelerations of these point masses and the forces exerted on them. It seems suitable to refer to it as a "three-dimensional diagram" rather than a "model," since it corresponds to a two-dimensional diagram of the velocity and acceleration of a point mass which is revolving in a circular path.

The gyroscope is thought of as constructed of a number of small masses attached to the outer ends of light rigid rods or spokes set radially into a shaft. To avoid the confusion of too many spokes in the field of view, the gyroscope is represented in the diagram (Fig. 1) by wooden disks of diameter 50 cm, to the faces of which are attached pieces of half-round stock to represent the spokes. Wooden balls of the sort used in the construction of crystal models may be attached to the rims of the disks, either at the ends of the spokes or at intermediate points,
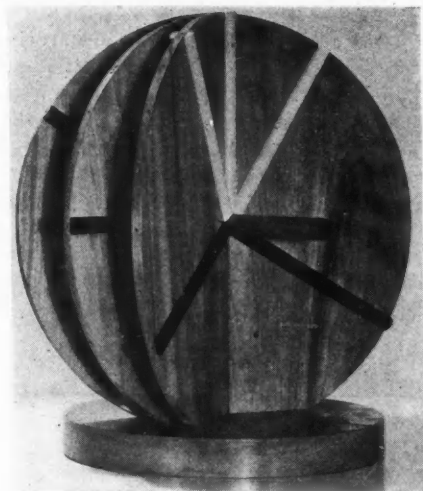
[1] Loeb, Am. Phys. Teacher 6, 66 (1938).

FIG. 1. Photograph of the 3-dimensional diagram, showing 3 successive positions of the plane of a precessing gyroscope.

by short dowels. The diagram shows three successive positions of a rotating and precessing gyroscope. The axis of precession (the vertical Z axis) is assumed to pass through the center of the gyroscope, or, if one prefers to think of the axis of precession as passing through one end of the axis of rotation as in the usual method of mounting, the three successive positions of the gyroscope have been displaced so as to bring the centers of the three disks into coincidence. The axis of rotation is shown in but one position, namely, perpendicular to the plane of the middle disk. One may think of the middle disk as representing the plane of the gyroscope "now," and the other disks, its position a moment ago and a moment hence.

A qualitative discussion of such a precessing point-mass system will be given first. Let us fix our attention on a single ball, the one that is "now" at the top of the gyroscope. Fig. 2 shows the position and velocity of this ball a moment ago. Fig. 3 shows the position and velocity of the ball "now." The spin velocity about the axis of rotation has displaced the ball through an angle θ, and the precessional velocity has displaced
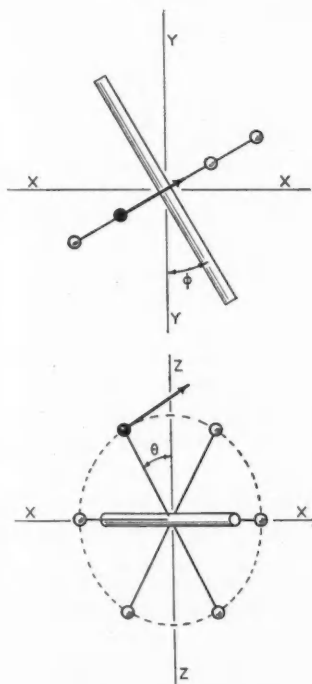

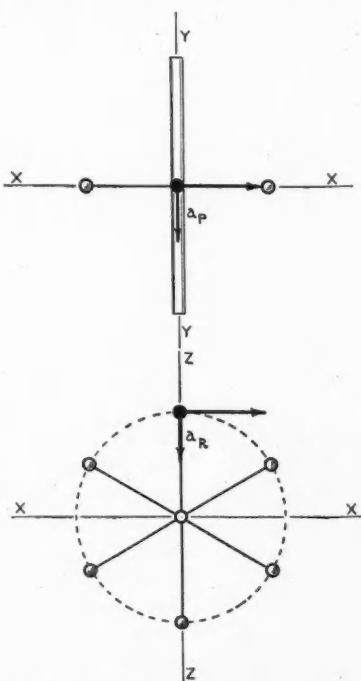FIG. 2. Position and velocity of a chosen ball a moment ago.

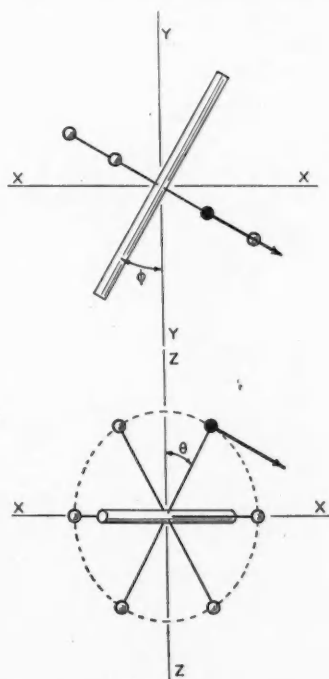FIG. 3. Position, velocity and accelerations of the same ball "now"

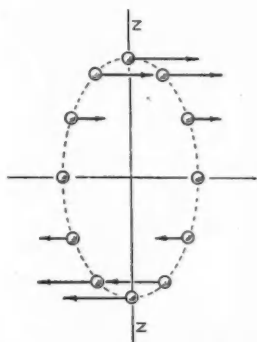FIG. 4. Position and velocity of the same ball a moment hence.

FIG. 5. Accelerations of a number of balls "now."

the axis through an angle ϕ. A moment later, the position and velocity of the same ball are as shown in Fig. 4.

If one looks at the three velocities along the Y axis, perpendicular to the $X-Z$ plane, it will be seen that there has been no change in the X component of the velocity between the initial and final positions. The Z component, however, has reversed from "up" to "down." Hence there has been an acceleration in the negative Z direction, or toward the center of the gyroscope. The average value of this acceleration, if the points are close together, is the acceleration of the ball "now," that is, at a point midway between the initial and final positions. This is the familiar radial acceleration (a_R, Fig. 3), to produce which it is necessary that the spoke shall be under tension.

If the three upper diagrams in Figs. 2 to 4 are compared, it will be seen that in moving

FIG. 6. Front view, showing velocity and precessional acceleration vectors.

from the initial to the final position the Y component of the velocity has also reversed, from "in" to "out." Hence, at the mid-point, the ball has a Y component of acceleration a_P because of the precessional motion. To produce this acceleration, the spoke to which the ball is attached must push sidewise on the ball. In other words, there will be a bending moment in the spoke in addition to a tension.
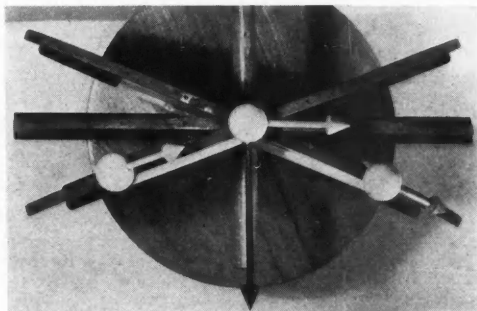

FIG. 7. Top view, showing velocity and precessional acceleration vectors.

If the changes in velocity of balls in positions other than at the top of the gyroscope are analyzed in the same way, it is easy to see (Fig. 5) that all those above the $X-Y$ plane are accelerated in the positive Y direction, and all those below the $X-Y$ plane are accelerated in the negative Y direction, the acceleration being a maximum for balls at the top and bottom, and decreasing to zero for those on the X axis.

If the number of balls is large, the positive and negative accelerations and, hence, the resultant Y forces will be equal in magnitude and opposite in direction, by symmetry. That is, in order to maintain the precessional motion, it is necessary that the spokes shall exert a couple on the balls, the sense of the couple being clockwise as one looks at the plane of the gyroscope from the left in Fig. 5. This couple can be provided only by an external couple acting on the gyroscope shaft.

In the usual method of mounting a precessing gyroscope, where one end of the spin axis rests on a fixed pivot, the couple is provided by the weight of the gyroscope and the upward thrust of the pivot. For, if the gyroscope axis is precessing in a horizontal plane, the resultant vertical force acting on the gyroscope must be
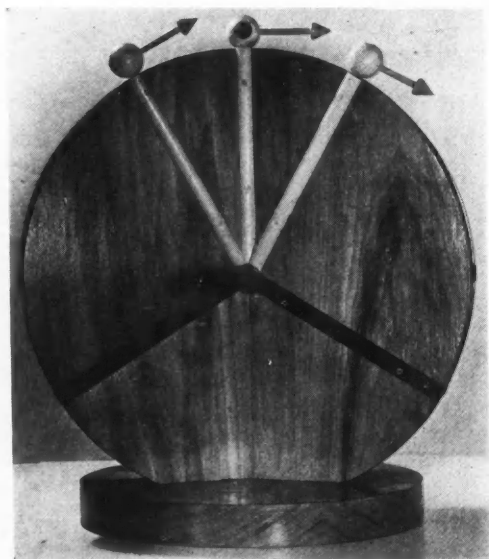
FIG. 8. Velocities and accelerations of a number of balls.

Figures 6 and 7 show front and top views of the diagram with three balls placed so as to find the accelerations at the top of the disk. (Only the precessional acceleration vector is shown.) Fig. 8 shows a number of balls in the "now" position, equipped with their velocity and acceleration vectors.

Instead of mounting the diagram on a base, as shown in the

FIG. 9. Accelerations of positive charges revolving in a magnetic field.

zero, and hence the upward thrust exerted by the pivot is equal in magnitude to the weight of the gyroscope.

All of the velocities and accelerations indicated in Figs. 2 to 5 can be readily illustrated with the aid of the three-dimensional diagram. In each of the wooden balls a number of radial holes are drilled in three mutually perpendicular directions. Wooden arrows to represent velocities and accelerations may be inserted in these holes, velocity vectors being painted one color and acceleration vectors another. The balls are inserted in sets of three at the rims of the three disks, in positions such that they represent three successive positions of the same ball. Each ball is provided with its velocity vector. By moving the diagram, or viewing it from the proper directions, it becomes evident at once just how the velocity is changing at any position of a ball. An acceleration vector of the proper magnitude and in the proper direction may then be inserted.

photographs, it would be preferable to support it from the ends of the axis of spin, so that the ball at the bottom of the disk might appear also.

This point-mass analysis of the precessing gyroscope may be carried over without change to the case of Larmor precession in a magnetic field. Fig. 9 represents a number of positive point charges revolving in a circular path in a uniform magnetic field whose direction is upward along the Z axis. From a straightforward application of the left-hand rule, the forces exerted on the charges will be seen to have the magnitudes and directions indicated in the figure. These forces produce accelerations that are precisely the same as those shown for the point masses in Fig. 8. The sidewise push, furnished by the spokes in the mechanical gyroscope, is in this case the side-thrust exerted on the charges by the magnetic field. Precession of the orbits takes place about the Z axis, or the direction of the magnetic field.
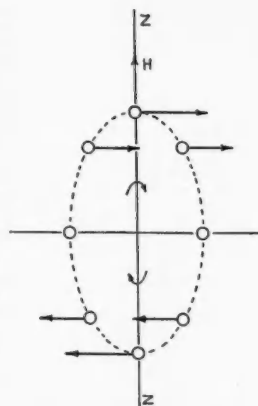
Attendance at the Stanford Meeting of the Association

The registration of those in attendance at the Stanford meeting of the American Association of Physics Teachers, June 28, 1939, lists 54 members and nonmembers. Members who registered are:

M. S. Allen, Long Beach Junior College; C. M. Bardin, Washington U. H. S., Freso; E. F. Barker, University of Michigan; R. T. Birge, University of California; B. F. Boardman, Stanford University; W. P. Boynton, Oregon State College; Edna Carter, Vassar College; A. E. Caswell, University of Oregon; N. W. Cummings, San Bernardino Junior College; L. E. Dodd, University of California at Los Angeles; W. W. Gayman, Sacramento Junior College; P. R. Gleason, Colgate University; L. W. Hartman, University of Nevada; F. E. Hoecker,

University of Kansas City; Mother Marie Kernaghan, Maryville College; H. A. Kirkpatrick, Occidental College; P. Kirkpatrick, Stanford University; A. A. Knowlton, Reed College; G. G. Kretschmar, Walla Walla College; V. F. Lenzen, University of California; K. V. Manning, University of Wichita; D. C. Miller, Case School of Applied Science; G. A. Pomeroy, San Mateo Junior College; A. Romer, Whittier College; G. O. Sagen, Bakersfield Junior College; R. S. Seward, College of Puget Sound; E. C. Watson, California Institute of Technology; D. L. Webster, Stanford University; Helen Wood, San Luis Obispo High School and Junior College.

An account of the meeting, with abstracts of contributed papers, appeared in the August issue, page 261.

DISCUSSION AND CORRESPONDENCE

Note on the Theory of the Magnetron

IN the theory of the cut-off relation for the magnetron[1] it is not necessary to consider all of the details of the electronic orbits. The expression for the potential difference V between the axial filament of radius a and the coaxial plate of radius b is

$$V = \frac{1}{8} \frac{e}{m} H^2 b^2 \left(1 - \frac{a^2}{b^2}\right)^2, \qquad (1)$$

if the magnetic field strength H is uniform. In the usual laboratory set-up for the determination of e/m by means of the magnetron, $a^2/b^2 \ll 1$, so that the error introduced by neglecting a^2/b^2 is small compared to errors from other sources. It is interesting to notice that with a^2/b^2 omitted, Eq. (1) is precisely the relation which is obtained if it is assumed that the electrons leave the axis of the tube with a radial velocity calculable from $Ve = \frac{1}{2}mv^2$ (where V equals the total difference of potential between the electrodes) and travel with constant speed in a space containing only the uniform magnetic field and no electric field; that is, travel in a perfect circle of radius $b/2$. It can be seen from qualitative considerations that the actual orbit of an electron which is just prevented from reaching the anode is a figure with bilateral symmetry and resembles a cardioid with its apex on the surface of the filament. For the foregoing case in which $a \ll b$, this figure must differ only slightly from a circle of radius $b/2$ passing through the axis. Evidently this must mean that with a thin filament the electric field in the vicinity of the filament is so intense that practically all of the potential energy of the electron has been changed to kinetic energy before it has gone an appreciable fraction of the radial distance between filament and anode. Stated mathematically this must mean that, as $a \to 0$,

$$\int_a^{a+\epsilon} dr/r \bigg/ \int_a^b dr/r \to 1 \qquad (2)$$

for any chosen number ϵ, no matter how small. That the ratio (2) does approach unity as a limit can be verified at once by writing it in the form

$$\lim_{a \to 0} \log\left(1 + \frac{\epsilon}{a}\right) \bigg/ \left(\log N + \log \frac{\epsilon}{a}\right),$$

where N is a number chosen so that $N\epsilon = b$. Since the ratio ϵ/a approaches infinity, whereas $\log N$ is finite, the equivalence of numerator and denominator in the limit is obvious.

LEONARD T. POCKMAN

Cornell University,
Ithaca, New York.

[1] See Page and Adams, *Principles of Electricity* (1932), p. 305.

Theory of the Triode as a Three-Body Problem in Electrostatics

IN an interesting paper Alexander Marcus[1] has discussed the electrostatic theory of the triode. In this connection it should be pointed out that the existence of space charge, so important for the operation of the triode as an amplifier, does not vitiate the voltage amplification factor derived on the basis of electrostatic theory. The electron current supplied by the cathode is normally limited by space charge, and under such conditions there is a definite functional relationship between this current and the field that would exist in the neighborhood of the cathode if no space charge were present. The electrostatic theory has shown that the grid is μ times as effective in producing field at the cathode as is the plate. Thus we might write

$$I = F\left(\frac{V_P}{\mu} + V_g\right), \qquad (1)$$

where F stands for a functional relationship of which the quantity in parentheses is the argument. Ordinarily the grid current can be neglected in comparison with the plate current; so Eq. (1) might be considered to represent the plate current. The voltage amplification factor, according to the ordinary definition, is then μ, regardless of the form of the function F.

Under practical circumstances, therefore, the amplification factor of a triode may be determined accurately from the interelectrode capacitances by taking the ratio of the grid-cathode capacitance to the plate-cathode capacitance. Max Abraham[2] has carried through the computation for a triode of coaxial cylindrical form. His cathode is a straight filament wire, the plate is a cylinder of radius r_p coaxial with it, and the grid consists of N straight wires of radius b, each parallel to the filament and equally spaced upon the geometrical surface of a cylinder of radius r_g coaxial with the plate and the filament and, of course, lying between them. In this case μ may ordinarily be written as

$$\mu = N \ln (r_p/r_g)/\ln (r_g/Nb). \qquad (2)$$

The writer has recently constructed such a triode, to be described elsewhere, and has found that the measured amplification factor checks that computed from Eq. (2) within 2 percent, which is well within the limits of experimental error. The interested reader will find a considerable literature on the subject of triode characteristics.[3]

PAUL L. COPELAND

Armour Institute of Technology,
Chicago, Illinois.

[1] Alexander Marcus, Am. Phys. Teacher 7, 196 (1939).
[2] Abraham, Archiv. f. Elektrotechnik 8, 42 (1919).
[3] For a logical presentation of the theory see Dow's *Fundamentals of Engineering Electronics* (Wiley, 1937). For discussions including results of experimental work see: Vodges and Elder, "Formulas for the Amplification Constant for Three Element Tubes," Phys. Rev. 24, 683 (1924); Kusunose, "Calculation of Characteristics and the Design of Triodes," Proc. I. R. E. 17, 1706 (1929).

Centrifugal Force and Circular Motion

In 1935 C. F. Hagenow[1] called attention to the fact that current textbooks contain definitions of centrifugal force which are "either incorrect, misleading or unnecessarily vague." He recommended that centrifugal force be defined as the reaction to centripetal force. In a recent paper, however, V. F. Lenzen[2] included the statement: "It would be desirable to restrict the term centrifugal force to the fictitious force in a rotating frame of reference." Statements made in elementary courses vary from a flat denial of the existence of a centrifugal force to its definition as a real, outward force of reaction equal in magnitude and opposite in direction to centripetal force.

The writer's purpose is to outline the inadequacies of customary treatments of circular motion at constant speed, and to describe in some detail a method by which the subject can be made understandable to beginners. Characteristic of the usual treatment is the tendency to avoid the term *centrifugal force* because of the confusion which ordinarily attends its use. The example of a particle constrained by centripetal force to motion in a circle is explained; one of the two well-known definitions of centrifugal force is given; and it is pointed out that centrifugal force need not be discussed further in the beginning course. Cream separator and centrifugal dryer are described in terms of centripetal force alone; other applications are given necessarily limited examination; and the study of circular motion is closed with an explanation designed to convince the students that particles leaving a rotating body do not "fly outward" in the usual sense.

Centrifugal force as the reaction of a body to an applied centripetal force is easily understood by the beginner if emphasis is placed on the fact that this reaction, since it is exerted *by* the given body rather than on it, does not affect the motion of the body. On the other hand, the fictitious centrifugal force (conceived as acting on the body to neutralize its centripetal force and enabling one to think of it as at rest in a frame of reference rotating with the body) is beyond the comprehension of the average beginner, although its use is an effective means of insuring that he will avoid the term *centrifugal force* and think in terms of centripetal force alone. At this point there is little choice between the two definitions: The first is more satisfying to the student in that it allows him to use the familiar term *centrifugal force* to designate a real, outward force; the second makes it perhaps less likely that he will experience confusion regarding centripetal force. It is interesting to note here that the fictitious centrifugal force, though defined as real only to observers in the rotating frame of reference, must be thought of as producing a stress real to observers in any frame of reference.

Pertinent objection to the usual treatment of circular motion in the elementary course[3] can be summarized by saying that it leaves the student in a confusion of half-understood examples and incomplete ideas. He reaches incorrect conclusions regarding outward forces; he cannot reconcile his ideas of tension, cohesion and reaction with his incomplete understanding of the applications of circular motion; he has no conception of the stresses set up within a rotating body. One illustration of the confusion caused by half-explained examples will be given here. In their attempts to demonstrate that a bursting flywheel simply "goes to pieces" instead of bursting in the usual sense, writers and instructors almost invariably lead students to the erroneous conclusion that no outward forces are set up within the rotating body. Typical of texts in which centrifugal force is either described in the fictitious sense or not defined at all is the following statement: "The term 'centrifugal force' is misleading in that it implies the existence of an outward force."[4] On the other hand, confusion often results from incomplete explanations based on the other definition of centrifugal force; for example, "Flywheels break because at high speeds the centrifugal force becomes very large."

Much of the confusion can be avoided through the careful discussion of certain examples that require the student to interpret the forces between the *parts* of a rotating system. A great deal can be gained by extending the standard example of a single particle to one in which two particles of masses M and m (Fig. 1) are constrained to circular motion with constant speed about a point O, M being on the line joining O and m, and M and m connected to O by a string of negligible mass. Let the force exerted on M by the string be F^i. Superscripts i and o will be used to designate forces as inward and outward, respectively, while subscripts M and m are to indicate association with masses M and m, respectively. If $C_M{}^i$ and $C_m{}^i$ are the forces that produce the centripetal accelerations of M and m, respectively, then $F^i = C_M{}^i + C_m{}^i$. It is evident that the forces on M must include, besides F^i, an outward force $C_m{}^o$, for the resultant (centripetal) force on M is, rearranging the first equation, $C_M{}^i = F^i - C_m{}^i = F^i + C_m{}^o = MA$, where A is the central acceleration of M. The force $C_m{}^o$ is recognized as the reaction exerted by the mass m and communicated to M by the string. For the present, let $C_m{}^o$ be called the *centrifugal reaction* of the mass m. Again, isolating the mass M, one observes that only F^i and $C_m{}^o$ act on it, and that their resultant is $C_M{}^i$, the centripetal force associated with M. A centrifugal reaction $C_M{}^o$ is exerted on the string by M, so that the combined centrifugal reaction is $C_m{}^o + C_M{}^o = F^o$, the outward force on the pivot at O. Conversely, the combined centripetal force
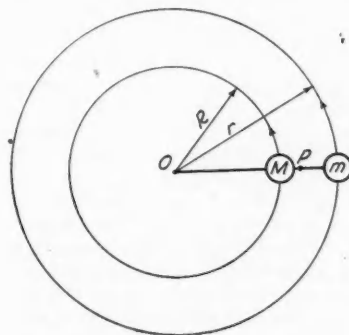

Fig. 1.

$F^i = C_M{}^i + C_m{}^i$ is exerted on the string by the pivot O. The *only* force acting on the outer mass m is the centripetal force $C_m{}^i$ communicated to it by the string. The emphasis indicated here by repetition and italics is typical of what is necessary to insure complete understanding; yet this emphasis is well spent, for the student who understands this example can be expected to retain and be able to apply correctly concepts of the forces set up in circular motion.

Now let us consider the conditions under which the string might break at a point P between M and m, where it is under a tension produced by the forces $C_m{}^i$ and $C_m{}^0$, which are of equal magnitude and opposite direction, and which are communicated to the string by M and m, respectively.[5] If it is supposed that the string at P cannot withstand a stretching force larger than T, the maximum angular speed which can be reached without breaking the string is that for which the foregoing forces each are of magnitude T. When this speed is exceeded the string parts, the cohesive forces (hence also centrifugal and central forces) fall off rapidly to zero, and m assumes a linear motion. The transition from circular to linear motion is thus a matter of the disappearance of the centripetal force and the acceleration it produces.

In the preceding example the term *centrifugal reaction* was introduced. In the writer's opinion the use of this term is of great advantage, for it satisfies the tendency of the student to associate the word *centrifugal* with an outward force, and at the same time it carries the reminder that the force is only a reaction. It makes unnecessary the use of the concept of fictitious centrifugal force, yet permits the definition of this concept without any ambiguity if the word *fictitious* is always included when reference to the fictitious force is intended. Again, the use of the term *centrifugal reaction* renders consistent the appearance of

the word *centrifugal* in connection with centrifugal pumps, driers, separators and the like. Finally, the self-descriptive character of the term facilitates concise explanations and leads naturally to a somewhat detailed examination of the usual examples. For example, it renders convenient the discussion of the distribution of stress within a rod rotating about one of its ends, the consideration of the distribution of pressure within a (simplified) centrifugal pump, and the explanation of why the speed of rotation required to dislodge small particles from the surface of a body is greater than for large ones. (Care should be taken to make clear whether the force here involved is adhesion or cohesion, as in the case of water drops.)

In the treatment of circular motion in the elementary course, then, the following steps are suggested as essential: (1) The usual correlation of centripetal force and central acceleration; (2) use of the term *centrifugal reaction* as described, and recognition of the term *centrifugal force* as identical in meaning but less desirable; (3) definition of *fictitious centrifugal force*, with the indication that it is useful as a mathematical tool in more advanced courses, and the statement that this fictitious or imaginary force is often confused with *centrifugal reaction;* (4) a careful treatment of the example of two revolving masses, and extension of the usual applied examples to include consideration of the stresses set up in rotation.

<div align="right">R. Orin Cornett</div>

Oklahoma Baptist University,
Shawnee, Oklahoma.

[1] C. F. Hagenow, Am. Phys. Teacher **3**, 190 (1935).
[2] V. F. Lenzen, Am. Phys. Teacher **7**, 66 (1939).
[3] The reference here is to the general type of treatment described earlier, *either* definition of centrifugal force being used.
[4] This statement is incorrect by virtue of incompleteness. Outward forces are set up *within* rotating bodies; what should be conveyed is that no outward forces act on the *outer* particles of a rotating system.
[5] It is assumed that the part of the string between O and M, which is under the larger tension, is strong enough not to break first.

The Pennsylvania Conference of College Physics Teachers

THE annual fall meeting of the Pennsylvania Conference of College Physics Teachers will be held at Pennsylvania State College on Friday and Saturday, October 20–21, 1939. The Friday afternoon session will consist as usual of contributed papers. The Friday evening and Saturday morning sessions will be devoted to the theme, "A Comparative Study of Agencies for Placing Physical Phenomena at the Disposal of the Student," and will include discussion of demonstration lectures, the conventional physics laboratory and the teaching museum. Special sessions for students will be held during the meeting, the students of Pennsylvania colleges being invited to contribute brief papers.

Lamme Medal for 1938

The Lamme Medal for 1938 has been awarded by the Society for the Promotion of Engineering Education to Dr. Stephen P. Timoshenko, Professor of Mechanics in Stanford University. The citation on the medal presented to Doctor Timoshenko reads: "for devotion to teaching as an art; for his distinguished classroom work as a teacher; for his ability to inspire his students toward ideals of scholarship and service; for his leadership in developing graduate work in mechanics, elasticity and strength of materials; for the enthusiasm he has imparted to many disciples to go from his classroom to devote themselves to research and to teaching; and for his continuing influence on them in their development." The Lamme Medal, founded to honor the memory of B. G. Lamme who at his death in 1924 was chief engineer of the Westinghouse Electric & Manufacturing Company, is presented annually to an engineer as recognition for outstanding accomplishment. Doctor Timoshenko is the twelfth medalist.

RECENT PUBLICATIONS

The Perception of Light. W. D. WRIGHT, Lecturer in Physics, Imperial College of Science and Technology, London. 105 p., 49 fig., 13×19 cm. *Blackie & Son*, 6s. net. This monograph emphasizes investigations in physiological optics that have an important bearing on technical problems of illumination and vision but that are likely to be overlooked by physicists and engineers who are working in these fields.

German-English Science Dictionary for Students in the Agricultural, Biological and Physical Sciences. LOUIS DE VRIES, Professor of Modern Languages, Iowa State College. 483 p., 13×18 cm. *McGraw-Hill*, $3. This carefully planned and excellently printed dictionary differs from other German-English dictionaries for science students in that it covers all the major sciences, including agriculture. It should be especially useful to the many workers in borderline fields of the sciences. Thirty-four members of the graduate faculty of Iowa State College collaborated with the author, physics being represented by WILLIAM KUNERTH.

Principles and Practices of Radio Servicing. H. J. HICKS, Radio Instructor, Vocational School, St. Louis. 315 p., 212 fig., 15×23 cm. *McGraw-Hill*, $3. Written especially for servicemen, this text first discusses the fundamental principles involved, using as little mathematics as possible, and then considers the different types of problems encountered in radio servicing. The suggestions and instructions for testing and servicing equipment are kept as general as possible but numerous diagrams and explanations are given of special circuits used by different radio manufacturers. Also included are explanations of the newer developments in radio receivers, such as automatic control of volume, tone and frequency, and useful suggestions concerning antenna installations for best reception.

High-Frequency Alternating Currents. KNOX MCILWAIN and J. G. BRAINERD, The Moore School of Electrical Engineering, University of Pennsylvania. Ed. 2. 542 p., 226 fig., 14×23 cm. *Wiley*, $6. Appearing originally in 1931, this mathematical analysis of the operation of electric circuits at high frequencies is designed to provide a comprehensive course of a year or more duration for senior or beginning graduate students in electrical engineering. Some familiarity with differential equations and with the elements of alternating currents is assumed. Problems and many specific references to the literature are included. In the present edition, new applications of the theory are given (for example, to high level modulators), the notation for vacuum tubes has been revised, and the Lagrange equation is developed and used in the discussion of electromechanical systems.

Astronomy. WILLIAM T. SKILLING, San Diego State College, and ROBERT S. RICHARDSON, Mount Wilson Observatory. 579 p., 133 fig., 116 plates, glossary, 14×22 cm. *Holt*, $3. The object of the authors of this textbook for a first course in astronomy is to present a picture of the science with better balance between things already known and things known only in part. Although the material is mainly that covered in a one-year course in a liberal-arts college, with rather more time devoted to arithmetic than is usual, it doubtless will find most favor in junior colleges and normal schools. The book work is excellent except that some plates and most of the tables would be more attractive if made somewhat larger. Certain rearrangements of the subject matter would improve its coherence: For example, the first chapter is devoted to the size, mass, age, rotation and revolution of the earth, and precession; 160 pages of material on coordinates, telescopes and solar radiation ensues; then, in Chapter VI, occurs the discussion of time, longitude and the calendar. It is not helpful to the student for diffuse nebulosities and exterior galaxies to be discussed in the same chapter. The authors are to be commended highly for the liberal use of photographs and diagrams, and for the inclusion of a glossary.

Theoretical Mechanics. C. J. COE, Assistant Professor of Mathematics, University of Michigan. 568 p., 102 fig., 14×21 cm. *Macmillan*, $5. Although this text makes consistent and excellent use of three-dimensional vector analysis, with the notation of Gibbs, its outlook and the selection of topics are in other respects the traditional ones of such writers as Appell, Routh and Whittaker. Mechanics is treated essentially as mathematics. In truth, the Preface of the book states that we may best "regard geometry, kinematics, classical mechanics, relativity mechanics as different theories, each resting on its own postulates, each consistent in itself and each possessing certain advantages in its degree of simplicity and in the degree of accuracy with which it interprets the physical world." In its recognition of this modern point of view the book is distinctly superior to most of the texts on mechanics that have been published for use in departments of mathematics. But the book is intended to provide a two-year course in classical mechanics, the first half being "introductory" to mechanics and obviously not assuming previous work in physics; and it is difficult to see how an introduction to physics through so-called "theoretical mechanics" can possibly give the student an understanding of, and feel for, genuine physics. In both spirit and practice, such courses in physical mathematics fail to emphasize and illustrate the continual interplay of experiment and theory that is the very essence of physics. Their strong point is mathematical rigor and elegance; their weak point, lack of physical rigor, which is a more general concept involving mathematical rigor as one necessary attribute.

DIGEST OF PERIODICAL LITERATURE

APPARATUS AND DEMONSTRATIONS

Convection currents. E. J. WILLIAMS; *Sch. Sci. Rev.* 20, 614, June, 1939. Shake thoroughly a can of cheap gold paint, pour the paint quickly into a bottle, then pour some of the paint into the shallow lid of the can and exhibit its surface on a screen by opaque projection. As the volatile liquid evaporates, thus cooling its upper layers, it divides into small cells in each of which the liquid ascends at the center, flows outward over the upper boundary and descends at the margin. The beauty of the demonstration is enhanced by the sheen of the gold.—D. R.

Safety device for a differential oil manometer. W. BARKAS; *J. Sci. Instr.* 16, 162, May, 1939. A differential U-tube oil manometer with a device for preventing the oil from blowing through tube when the pressures are not approximately equal is shown in Fig. 1. The bottom of the

FIG. 1. Safety device for an oil manometer.

U-tube, of wider diameter than the rest, contains a steel ball bearing. A sudden difference of pressure drives this ball against one of the narrower ends and closes the tube; with nearly equal pressures the ball, rolling to the center, does not affect the equilibrium. Grinding of the constricted necks is unnecessary unless the pressure difference is to be held for long periods. If the ball sticks after the pressures are equalized, it may be loosened by tapping or with a magnet.—H. N. O.

INDUSTRY AND "HIGHBROW" PHYSICS

Industry's quarrel with physics. ANON; *J. Sci. Instr.* 16, 238, July, 1939. Because of its "special interest to those responsible for the training of physicists for British industries," the following editorial which appeared in the May issue of the American journal *Metals and Alloys* has been reprinted in full by the *Journal of Scientific Instruments* (London).

"Industry's chief quarrel with physics, as it has been taught in the last decade, is that it tends to over-emphasize the quantum point of view and . . . speculations on atomic structure to the exclusion of more tangible problems. The tendency has been . . . to train men to be professors of highbrow physics that they may train other men to be, in their turn, professors of a similar type. Inbreeding, over-production and gross neglect of everyday engineering physics has resulted. It is a pleasure to note a movement within the ranks of the physicists themselves to remedy this situation.

"A physicist with his feet on the ground is a most useful animal. One with his head in the clouds may be potentially useful in respect to pure science, but the proportion of physicists of sufficient stature to have both extremities in both places is very small. We'd rather see all physicists trained so thoroughly in truly engineering physics that all will have their feet on the ground and only those inherently of sufficient stature get their heads in the clouds. Only heads of men of large stature can look into the future anyhow. Many who are incompetent in respect to forward-looking pure physics would be competent to apply known physics to engineering and to broaden the base of engineering by research in lowbrow physics.

"In the last decade or so, many men have been permanently spoiled by their graduate work in physics of a type for which they are not fitted, but which type they nevertheless stick to, or remain unemployed because of the lack of positions in work requiring that brand of physics. They have not had adequate grounding in the type of physics that industry needs, and could profitably create jobs in. If engineers could be induced to take graduate work in engineering physics after their regular course in engineering, or if graduates in general physics did their own graduate work in engineering rather than in highbrow physics, the situation might be improved. At any rate, the more discussions that are held on the down-to-earth type of physics, the better the chance for much needed improvement in a situation that vitally concerns the future of metallurgical engineering."

This editorial appeared in *Metals and Alloys* in connection with an announcement concerning the symposium on *Temperature and Its Measurement in Science and Industry* which will be held in New York on November 2–4 under the sponsorship of the American Institute of Physics.—D. R.